Renal cell cancer

Renal Cell Cancer:
an epidemiological approach to unravel disease heterogeneity

Jeroen van de Pol

Renal Cell Cancer: an epidemiological approach to unravel disease heterogeneity

ISBN: 978-94-6416-355-1
Design Cover: Jeroen van de Pol
Derivative of 'Kofe illuziya3' by Berserkerus - (CC BY-SA 2.5)
Layout: Jeroen van de Pol
Print: Ridderprint | www.ridderprint.nl

The work presented in this thesis was performed within GROW, School for Oncology and Developmental Biology at the Department of Epidemiology, Maastricht University, The Netherlands. The analyses in this dissertation are based on data from the Netherlands Cohort Study on Diet and Cancer (NLCS).
Printing and dissemination of this thesis was financially supported by the Department of Epidemiology at Maastricht University.

Renal Cell Cancer:
an epidemiological approach to unravel disease heterogeneity

DISSERTATION

To obtain the degree of Doctor at Maastricht University, on the authority of the Rector Magnificus, Prof. Dr. Rianne M. Letschert in accordance with the decision of the Board of Deans, to be defended in public on
Thursday, 11th of February 2021 at 16:00 hours

by

Jeroen Albert Adrianus van de Pol
Born 29th of October 1990, Eindhoven, the Netherlands

Promotor

Prof. Dr. Ir. P.A. van den Brandt

Co-Promotor

Dr. L.J. Schouten
Dr. K. Kok (UMCG Groningen)

Assessment Committee

Prof. dr. J.P. Kooman (*Chair*)
Prof. dr. A. zur Hausen
Prof. dr. L.A.L.M. Kiemeney (Radboud UMC)
Prof. dr. P.F.A. Mulders (Radboud UMC)
Prof. dr. F.C.S. Ramaekers

Table of contents

CHAPTER 1

General introduction

Worldwide, over 403.000 individuals were diagnosed with kidney cancer in 2018[1]. This accounts for 2-3% of all newly diagnosed cancers, which makes kidney cancer the 14th most common cancer in the world. The incidence of kidney cancer is twice as high in men as in women[2, 3]. In addition, the incidence increases steadily with age, with the peak incidence at approximately 75 years[2]. The age-standardised incidence rate (according to the World standard population; ASR) of kidney cancer is reported to be particularly high in more developed regions (9.4 per 100.000), including The Netherlands, when compared to less developed regions (ASR <1 per 100,000)[1].

More than 90% of malignancies in the kidney are comprised of carcinoma of the renal parenchyma or renal cell carcinoma (RCC). The ASR of RCC in The Netherlands has been increasing steadily from 6.0 per 100,000 in 1989 to 7.3 per 100,000 in 2007 and has since stabilised (7.3 per 100,000 in 2017)[4]. The initial rise in incidence has largely been attributed to the more widespread use and the improvement of imaging techniques such as computed tomography (CT) and ultrasonography[5-7].

Tumour histology

Renal cell carcinoma can be classified into several histopathological entities with distinct pathological and genetic characteristics. Clear cell RCC (ccRCC; 70-80%) and papillary RCC (pRCC; 10-17%) account for the majority of all RCC occurrences[8]. The other RCCs are comprised of chromophobe RCC (chRCC; 5-8%), and various other less frequent subtypes (<1%)[8, 9].

Clear cell RCC is thought to arise from epithelial cells in the proximal convoluted renal tubule and are characterised by their optically clear cytoplasm[10]. Clinically, the prognosis of ccRCC is largely dependent on the clinical characteristics at diagnosis. This is often exemplified by the TNM Classification of Malignant Tumours (TNM), in which information on tumour size and extension (T), lymph node involvement (N) and distant metastasis (M) can be used to classify tumours by anatomical stage. Stage I ccRCC has a 5-year survival of 86%, while stage IV ccRCC has a 5-year survival of 18%[11]. In addition, other features such as tumour grade, necrosis, sarcomatoid/rhabdoid differentiation and microvascular invasion can be used as prognostic parameters to predict the cancer-specific survival of patients[12].

Papillary RCC is thought to have a better 5-year survival rate and, in turn, a lower malignant potential, when compared to ccRCC[10, 13, 14]. Papillary renal cell carcinoma is generally subdivided into two morphologically distinct subtypes, type 1 and type 2 pRCC. Type 1 pRCC is often multifocal and further characterised by papillae and tubular structures, covered with small cells containing pale cytoplasm and small, uniform, oval nuclei[15]. Type 2 pRCC is characterised by papillae covered with large cells containing eosinophilic cytoplasm and large, spherical nuclei[15]. Clinically, type 1 pRCC generally presents with a lower tumour grade and stage compared to type 2 pRCC, and is thought to have a better overall survival and a less aggressive nature[14-16]. However, understanding the prognostic implications of the distinction between type 1 and type 2 pRCC remains challenging as results on the prognosis of pRCC subtypes have been inconsistent, which is largely attributed to confounding by clinical characteristics[17].

Tumour genetics of ccRCC

Cancer is caused by the gradual accumulation of DNA damage in an individual cell during a person's life. As such, these particular mutations are not transmitted from one generation to the other, but transmitted to all cells descending from the mutated cell[18]. This provides cancer researchers the opportunity to use DNA mutations as a marker to identify the steps that contributed to the development of cancer. Mutations in driver genes may lead to the reprogramming of cell growth, the breakdown of genomic maintenance systems and the reprogramming of cell metabolism, which may result in abnormal cell growth[19].

The most characteristic features of ccRCC are the functional loss of *VHL*, and the loss of the short arm of chromosome 3 (3p), with an incidence of over 90% in primary ccRCC[20]. *VHL* was originally identified through studies on the hereditary von Hippel-Lindau syndrome[21]. In this syndrome, the inactivation of the remaining copy of *VHL* through somatic mutations, gene silencing or deletion, in addition to an already present germline mutation of the other allele, was discovered as a characteristic factor in the development of ccRCC[21]. Due to the presence of a germline *VHL* mutation, only a *VHL* mutation in the remaining allele in any of the proximal epithelial cells of the kidney suffices to initiate the development of ccRCC. Thus, 90% of the individuals with the von Hippel-Lindau syndrome tend to develop renal tumours before the age of 65, with a mean age of RCC development of approximately 40 years[22-24]. Mutations in *VHL* have also been implicated in sporadic ccRCC[25]. In keeping with Knudson's two-hit hypothesis, sporadic RCC requires the inactivation of both parental *VHL* alleles[26-28]. However, not all individuals with sporadic ccRCC have a mutation in *VHL*, which suggests that involvement of other genes may also be involved in the aetiology of ccRCC[8]. Indeed, other tumour suppressor genes present on chromosome 3p besides *VHL*, namely *PBRM1, BAP1* and *SETD2,* have been implicated in the development of ccRCC[15]. Furthermore, several other genetic alterations have been detected in various pathways, including the VHL/HIF-, and mTORC1-pathways and the SWI/SNF-complex[29-31].

Risk factors

Environmental exposures are thought to affect the somatic mutation rate[32]. By reducing the exposure to environmental factors that may enhance the mutation rate the probability that driver genes will become mutated may be reduced[32]. Therefore, identifying (modifiable) risk factors poses an opportunity for primary prevention strategies against cancer.

A multitude of modifiable risk factors has been investigated in relation to the overall risk of RCC. However, few risk factors have been consistently associated with RCC. Established modifiable risk factors for RCC include cigarette smoking, obesity and hypertension[2, 31]. Interestingly, while alcohol consumption is generally thought to increase cancer risk, it has been inversely associated with the risk of RCC[31, 33]. Overall, the effect estimates associated with these risk factors are often modest. However, due to the high prevalence of these risk factors in the population, the attributable risks are often sizable[34].

Large meta-analyses evaluating the role of tobacco and cigarette smoke exposure have concluded that there is a clear indication of an association with RCC risk. Evidence from a meta-analysis, including evidence from 38 case-control studies and 22 cohort studies, indicated an increased risk in both former smokers (pooled RR 1.16 (95%CI 1.08-1.25) and

current smokers 1.36 (1.19-1.56), when compared to non-smokers[35]. An earlier meta-analysis reported a strong dose-dependency in the RCC risk increase in both men and women, as well as a reduction in RCC risk with increasing duration of smoking cessation[36].

Being overweight or obese has also been indicated as major risk factor for RCC as well. Evidence from a dose-response meta-analysis, combining information from 21 cohort studies, indicated that overweight (Body Mass Index (BMI); between 25 and <30 kg/m^2) was associated with an increased risk of developing RCC (pooled RR 1.28 (95%CI 1.24-1.33)), while the association of obesity (BMI ≥ 30 kg/m^2) with the risk of RCC was stronger (1.77 (95%CI 1.68-1.87), when compared to normal weight[37].

There is a large body of evidence regarding the association between having a history of hypertension and the risk of RCC, although some controversy exists regarding the role of antihypertensive medication in this relationship[38]. Based on findings from a meta-analysis, which included 12 prospective studies, a positive association between hypertension and the risk of RCC was reported (pooled RR 1.67 (95%CI 1.46-1.90)[38]. Multiple studies have attempted to disentangle the effect of antihypertensive medication from the effect of hypertension[39]. One study reported an association between hypertension and kidney cancer, independent of antihypertensive use[39]. Another study reported that a positive association of antihypertensive use with RCC risk was only detected in combination with poorly controlled blood pressure[39]. Whereas, one study reported an increased risk in individuals who used antihypertensive medication[40], others indicated that the effect of antihypertensive medication disappeared after correction for hypertension[41, 42]. These discrepancies highlight that there are still some knowledge gaps regarding the relationship between hypertension and the risk of developing RCC.

In strong contrast to most cancers, moderate alcohol consumption has been reported to be inversely associated with the risk of developing RCC (pooled RR 0.79 (95%CI 0.72-0.86))[33]. However, no further benefit was attained for levels above moderate alcohol consumption[33]. In fact, one meta-analysis reported that the beneficial reduction in RCC risk may already be attained at one alcoholic beverage per day[43]. Although the mechanisms behind this association are poorly understood some hypotheses regarding the effects of alcohol on insulin sensitivity, or its diuretic effect have emerged in the literature[33, 44, 45].
While these factors have been associated in a plethora of epidemiological studies, the evidence for multiple other risk factors from prospective cohort studies is limited or lacking. As a result, many risk factors and their role in RCC carcinogenesis remain poorly understood.

Aside from these established and consistently associated risk factors multiple other potential risk factors have been implicated in the aetiology of RCC. As a result, potential opportunities for the prevention of RCC remain underrepresented in scientific literature. Among those risk factors are type 2 diabetes mellitus, kidney stones, physical activity and dietary factors, e.g. the intake of vegetables and fruits. The knowledge gap resides partially in the lack of extensive prospective cohort studies that have assessed these factors in relation to RCC, and partly in the difficulty of adequately assessing these risk factors. Evidence from large-scale prospective cohort studies on these potential risk factors may, therefore, provide additional knowledge on the aetiology of RCC, which may help with further uncovering the biological

mechanisms driving this cancer.

Aetiologic heterogeneity

In recent years, the aetiology of renal cell carcinoma is more often analysed by stratifying by histological subtype. Based on the aforementioned distinct clinical, pathological and genetic differences, it has been speculated that the RCC subtypes may possess distinct aetiologies. While the current evidence remains sparse, there are some indications for such aetiologic heterogeneity across RCC subtypes. The most prominent heterogeneity has been observed for BMI. Current evidence suggests a positive association between BMI and ccRCC, but no association was observed with pRCC[46, 47]. Furthermore, differences in the prevalence of ccRCC and pRCC have been found in current smokers[48]. However, other studies observed no etiologic heterogeneity regarding cigarette smoking[46, 47]. Moreover, based on the currently available scientific literature, no clear etiologic heterogeneity has been indicated for history of hypertension across ccRCC and pRCC risk[46, 47]. While no heterogeneity has been observed for hypertension, a potential heterogeneity exists regarding the use of antihypertensives[49]. In particular, pRCC was associated with long-term diuretics use and calcium channel blockers, while the reported associations for ccRCC were weaker[49]. These findings highlight the need for further research to bolster the evidence base for future preventative measures against RCC.

Molecular Epidemiology and RCC

Molecular epidemiology aims to establish the relationship between biomarkers on the one hand, and exposures, susceptibility and disease on the other[18]. In RCC, multiple molecular methods are being employed to learn more about the aetiologic makeup of RCC and its subtypes, and the mechanisms that drive the occurrence or prognosis of RCC. In this thesis, we have used information on germline variants, in addition to information on environmental exposures, to investigate gene-environment interactions and ccRCC risk.

Germline single nucleotide polymorphisms (SNPs)

As stated before, epidemiologic observations have implicated environmental factors in the development of cancer[50]. However, the exposure to environmental factors can not entirely explain cancer risk, indicating that there is a crucial role of genetic susceptibility among similarly exposed individuals[50]. This notion has led to the use of gene-environment interaction approaches, in which epidemiologists try to obtain better estimates of disease risks by accounting for the joint effect of genes and environmental exposures[51, 52]. One way to study this is by using single nucleotide polymorphisms (SNPs). SNPs are commonly occurring genetic variations present in the germline. When SNPs are present in a regulatory region of a gene they may affect the gene's function or its expression level, which may lead to increased disease susceptibility[51]. The effects of these SNPs may be amplified upon interaction with a specific environmental exposure. These gene-environment interactions play a role in a large fraction of cancer cases in a population, due to the high frequency of their occurrence[50]. In (cc)RCC, multiple gene-environment interactions have already been identified that convey additional risk. For instance, such interactions have been observed for (cc)RCC risk regarding *ADH7* and alcohol consumption[53]; *AGTR*, *AGT* and *ACE* and sodium and hypertension[54]; *RXRA* and calcium and vitamin D intake[55]; *NAT2*, *CYP1A1* and *GSTM1* and tobacco smoking[56]; and *ITPR2* and *EPAS1* and meat-cooking mutagens[57]. With

the continuous detection of new susceptibility loci, more gene-environment interactions can be expected to be found. This may potentially lead to more insight in the population-based differences in cancer risk and into the mechanisms of carcinogenesis.

Somatic mutations

Large efforts have been made using next-generation sequencing techniques to characterize the genomic landscape of RCC of which the data is widely accessible using large open databases, such as The Cancer Genome Atlas (TCGA), the Catalogue of Somatic Mutations in Cancer (COSMIC) database and cBioportal[16, 58-60].

Through these efforts, a wealth of information is available on, among others, the occurrence of somatic mutations in ccRCC. Data generated by these large scientific collaborations indicates that *VHL* is the most frequently mutated gene in ccRCC (~52% mutated)[29, 58]. In addition to *VHL*, somatic mutations in other genes have also been implicated in ccRCC, namely *PBRM1* (30-33%), *SETD2* (12-13%), *BAP1* (10-13%), *KDM5C* (7%), *MTOR* (6-7%), *TP53* (2-7%), and multiple less frequently mutated genes (<5%)[29, 58]. There are indications that somatic mutations also play a role in the prognosis of (cc)RCC[61-66]. For instance, somatic mutations in *PBRM1* have been associated with a better overall survival for ccRCC[66]. Moreover, mutations in *BAP1, SETD2, KDM5C* and *TP53* have been associated with an unfavourable prognosis of ccRCC[61-63].

The implications of the use of large databases in ccRCC research

There is a need for future research regarding the involvement of various mutations and pathways in ccRCC as there are some limitations involved in the use of large databases, such as the TCGA and COSMIC. For instance, information available in the TCGA is assembled mainly from a convenience sample of cancer patients, and although strict criteria were in place for eligibility, there appear to be differences in the general population with regards to age, race/ethnicity and clinical characteristics for the majority of cancers[67]. For renal cell carcinoma, an overrepresentation of stage 3 and 4 tumours and an underrepresentation of stage 1 tumours was observed, when compared to the U.S. population. This is likely related to the stringent inclusion criteria[67]. In addition, TCGA samples are primarily provided by U.S. academic institutions and, in part, derived from research and trials in these academic centres[68, 69]. As a result, this may limit the generalisability of findings to all patients as individuals treated in these institutions are generally younger, compared to the average cancer patient[67]. As the COSMIC database is largely built upon the sample information provided by the TCGA, a lot of the current scientific knowledge is driven by results from the TCGA[58]. Lastly, the TCGA was not originally designed to study risk factors or to perform survival analyses. Consequently, the completeness and accuracy of the information available on these factors is lower than in traditional epidemiologic studies and may warrant additional caution when interpreting results. Therefore, it is of great importance to further the current knowledge by using information from population-based prospective cohort studies with a long follow-up to obtain an accurate representation of the genomic landscape of cancer cases at the population-level.

In conclusion, the many aetiological mechanisms present in the development of RCC make it a complex disease to analyse. Furthermore, evidence from large prospective cohort studies regarding the heterogeneity of associations between risk factors and the development of

histologic subtypes is currently lacking. Additional insight into the interplay of environmental factors and the genetic make-up may open up avenues for future research into the mechanisms driving the development of this cancer.

Rationale of this thesis

Using information from a large prospective cohort study we aimed to identify whether various environmental and genetic risk factors are differentially associated with the risk of specific subgroups of renal cell carcinoma. In addition, we investigated whether somatic mutations were associated with the prognosis of clear cell renal cell carcinoma.

Study design

All studies within this thesis were conducted using information obtained through the NLCS. The NLCS was initiated in September 1986 with the inclusion of 120,852 men and women aged 55-69 years from 204 Dutch municipal population registries[70]. At baseline, all participants completed a mailed, self-administered questionnaire on dietary habits and other risk factors for cancer, including lifestyle factors, medical conditions and anthropometry. The questionnaire included a 150-item food frequency questionnaire that focused on habitual food consumption during the year preceding baseline. In addition to the baseline questionnaire, approximately 90,000 participants provided toenail clippings, which have shown to be a valid source of DNA for the genotyping of germline genetic variants[71].

In the NLCS, a case-cohort design was used for efficiency in data processing, genotyping and follow-up for vital status. Cases were derived from the entire cohort, while a randomly selected subcohort of 5,000 participants was used to estimate person-time for the entire cohort[70]. Follow-up for cancer occurrence for all participants was conducted by computerised record linkage with the Netherlands Cancer Registry (NCR), the Dutch pathology registry (PALGA), and the causes of death registry maintained by Statistics Netherlands (CBS). The follow-up for vital status of the subcohort was nearly 100% complete after 20.3 years. The completeness of cancer follow-up is estimated to be over 96%[72].

During the 20.3 years of follow-up, 608 RCC cases were identified. Cases with histologically confirmed epithelial RCC were eligible for the collection of formalin-fixed paraffin-embedded (FFPE) tumour tissue. In total 454 FFPE tumour blocks were collected from ~50 pathology laboratories throughout the Netherlands. Two experienced pathologists revised the tumour histology according to the WHO-classification of RCC tumours[73]. Further information on the clinical characteristics, such as age at diagnosis and tumour size, were obtained from pathological reports and the Netherlands cancer registry. Information on survival time was accomplished by record linkage with municipal population registries and the causes of death registry maintained by Statistics Netherlands (CBS). The identification of somatic mutations was performed using a targeted sequencing protocol for a panel of 42 genes implicated in ccRCC on DNA isolated from FFPE tumour blocks of 252 cases.

Thesis outline

In **chapter 2** we describe the association of type 2 diabetes and its treatment with renal cell cancer risk. In **chapter 3 and 4** we assessed the heterogeneity of associations across ccRCC and pRCC for various risk factors for RCC. **Chapter 3** details analyses on the association

between established risk factors for RCC and histologic subtypes of RCC, including cigarette smoking, body mass index, alcohol consumption and history of hypertension. **Chapter 4** details the association between kidney stones and ccRCC and pRCC risk. In addition, we describe the association between kidney stones at various localisations of upper tract urothelial carcinoma. In **chapter 5**, we studied the association between four germline polymorphisms on (cc)RCC risk. In addition, we assessed potential gene-environment interactions, gene-gene interactions and the association between SNPs and *VHL* promoter methylation status. In **chapter 6**, we created a seven-gene mutational profile and assessed whether somatic alterations in these genes affected the cause-specific survival of RCC, taking into account patient characteristics and the co-occurrence of mutations. Lastly, in **chapter 7**, we discuss the implications of our results in light of the current literature and discuss future perspectives of research into RCC.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68: 394-424.
2. Scelo G, Larose TL. Epidemiology and Risk Factors for Kidney Cancer. J Clin Oncol. 2018: Jco2018791905.
3. Znaor A, Lortet-Tieulent J, Laversanne M, Jemal A, Bray F. International variations and trends in renal cell carcinoma incidence and mortality. Eur Urol. 2015;67: 519-530.
4. Nederlandse Kankerregistratie (NKR), IKNL. NCR data & figures, iknl.nl/nkr-cijfers, 13 May 2020.
5. Chow WH, Devesa SS, Warren JL, Fraumeni JF, Jr. Rising incidence of renal cell cancer in the United States. Jama. 1999;281: 1628-1631.
6. Hock LM, Lynch J, Balaji KC. Increasing incidence of all stages of kidney cancer in the last 2 decades in the United States: an analysis of surveillance, epidemiology and end results program data. J Urol. 2002;167: 57-60.
7. Hollingsworth JM, Miller DC, Daignault S, Hollenbeck BK. Rising incidence of small renal masses: a need to reassess treatment effect. J Natl Cancer Inst. 2006;98: 1331-1334.
8. Baldewijns MM, van Vlodrop IJ, Schouten LJ, Soetekouw PM, de Bruine AP, van Engeland M. Genetics and epigenetics of renal cell cancer. Biochim Biophys Acta. 2008;1785: 133-155.
9. Truong LD, Shen SS. Immunohistochemical diagnosis of renal neoplasms. Arch Pathol Lab Med. 2011;135: 92-109.
10. Cheville JC, Lohse CM, Zincke H, Weaver AL, Blute ML. Comparisons of outcome and prognostic features among histologic subtypes of renal cell carcinoma. Am J Surg Pathol. 2003;27: 612-624.
11. Feng X, Zhang L, Tu W, Cang S. Frequency, incidence and survival outcomes of clear cell renal cell carcinoma in the United States from 1973 to 2014: A SEER-based analysis. Medicine (Baltimore). 2019;98: e16684.
12. Delahunt B, Cheville JC, Martignoni G, et al. The International Society of Urological Pathology (ISUP) grading system for renal cell carcinoma and other prognostic parameters. Am J Surg Pathol. 2013;37: 1490-1504.
13. Amin MB, Corless CL, Renshaw AA, Tickoo SK, Kubus J, Schultz DS. Papillary (chromophil) renal cell carcinoma: histomorphologic characteristics and evaluation of conventional pathologic prognostic parameters in 62 cases. Am J Surg Pathol. 1997;21: 621-635.
14. Delahunt B, Eble JN, McCredie MR, Bethwaite PB, Stewart JH, Bilous AM. Morphologic typing of papillary renal cell carcinoma: comparison of growth kinetics and patient survival in 66 cases. Hum Pathol. 2001;32: 590-595.
15. Linehan WM, Spellman PT, Ricketts CJ, et al. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. N Engl J Med. 2016;374: 135-145.
16. Ricketts CJ, De Cubas AA, Fan H, et al. The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. Cell Rep. 2018;23: 313-326.e315.
17. Wong ECL, Di Lena R, Breau RH, et al. Morphologic subtyping as a prognostic predictor for survival in papillary renal cell carcinoma: Type 1 vs. type 2. Urol Oncol. 2019;37: 721-726.
18. Rothman N, Hainaut P, Schulte P, Smith M, Boffetta P, Perera F. Molecular Epidemiology Principles and Practices. Lyon, France: IARC Scientific Publications, No. 163, 2011.
19. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144: 646-674.
20. Hsieh JJ, Le VH, Oyama T, Ricketts CJ, Ho TH, Cheng EH. Chromosome 3p Loss-Orchestrated

VHL, HIF, and Epigenetic Deregulation in Clear Cell Renal Cell Carcinoma. J Clin Oncol. 2018;36: Jco2018792549.
21. Latif F, Tory K, Gnarra J, et al. Identification of the von Hippel-Lindau disease tumor suppressor gene. Science. 1993;260: 1317-1320.
22. Haas NB, Nathanson KL. Hereditary kidney cancer syndromes. Adv Chronic Kidney Dis. 2014;21: 81-90.
23. Ong KR, Woodward ER, Killick P, Lim C, Macdonald F, Maher ER. Genotype-phenotype correlations in von Hippel-Lindau disease. Hum Mutat. 2007;28: 143-149.
24. Roupret M, Hopirtean V, Mejean A, et al. Nephron sparing surgery for renal cell carcinoma and von Hippel-Lindau's disease: a single center experience. J Urol. 2003;170: 1752-1755.
25. Gnarra JR, Glenn GM, Latif F, et al. Molecular genetic studies of sporadic and familial renal cell carcinoma. Urol Clin North Am. 1993;20: 207-216.
26. Kaelin WG. Von Hippel-Lindau disease. Annu Rev Pathol. 2007;2: 145-173.
27. Kim WY, Kaelin WG. Role of VHL gene mutation in human cancer. J Clin Oncol. 2004;22: 4991-5004.
28. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A. 1971;68: 820-823.
29. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013;499: 43-49.
30. Varela I, Tarpey P, Raine K, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature. 2011;469: 539-542.
31. Hsieh JJ, Purdue MP, Signoretti S, et al. Renal cell carcinoma. Nat Rev Dis Primers. 2017;3: 17009.
32. Tomasetti C, Li L, Vogelstein B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. Science (New York, N.Y.). 2017;355: 1330-1334.
33. Bagnardi V, Rota M, Botteri E, et al. Alcohol consumption and site-specific cancer risk: a comprehensive dose-response meta-analysis. Br J Cancer. 2015;112: 580-593.
34. Benichou J, Chow WH, McLaughlin JK, Mandel JS, Fraumeni JF, Jr. Population attributable risk of renal cell cancer in Minnesota. Am J Epidemiol. 1998;148: 424-430.
35. Cumberbatch MG, Rota M, Catto JW, La Vecchia C. The Role of Tobacco Smoke in Bladder and Kidney Carcinogenesis: A Comparison of Exposures and Meta-analysis of Incidence and Mortality Risks. Eur Urol. 2016;70: 458-466.
36. Hunt JD, van der Hel OL, McMillan GP, Boffetta P, Brennan P. Renal cell carcinoma in relation to cigarette smoking: meta-analysis of 24 studies. Int J Cancer. 2005;114: 101-108.
37. Wang F, Xu Y. Body mass index and risk of renal cell cancer: a dose-response meta-analysis of published cohort studies. Int J Cancer. 2014;135: 1673-1686.
38. Hidayat K, Du X, Zou SY, Shi BM. Blood pressure and kidney cancer risk: meta-analysis of prospective studies. J Hypertens. 2017;35: 1333-1344.
39. Weikert S, Boeing H, Pischon T, et al. Blood pressure and risk of renal cell carcinoma in the European prospective investigation into cancer and nutrition. Am J Epidemiol. 2008;167: 438-446.
40. Sanfilippo KM, McTigue KM, Fidler CJ, et al. Hypertension and obesity and the risk of kidney cancer in 2 large cohorts of US men and women. Hypertension. 2014;63: 934-941.
41. Shapiro JA, Williams MA, Weiss NS, Stergachis A, LaCroix AZ, Barlow WE. Hypertension, antihypertensive medication use, and risk of renal cell carcinoma. Am J Epidemiol. 1999;149: 521-530.

42. Yuan JM, Castelao JE, Gago-Dominguez M, Ross RK, Yu MC. Hypertension, obesity and their medications in relation to renal cell carcinoma. Br J Cancer. 1998;77: 1508-1513.
43. Song DY, Song S, Song Y, Lee JE. Alcohol intake and renal cell cancer risk: a meta-analysis. Br J Cancer. 2012;106: 1881-1890.
44. Wozniak MB, Brennan P, Brenner DR, et al. Alcohol consumption and the risk of renal cancers in the European prospective investigation into cancer and nutrition (EPIC). Int J Cancer. 2015;137: 1953-1966.
45. Altieri A, La Vecchia C, Negri E. Fluid intake and risk of bladder and other cancers. Eur J Clin Nutr. 2003;57 Suppl 2: S59-68.
46. Purdue MP, Moore LE, Merino MJ, et al. An investigation of risk factors for renal cell carcinoma by histologic subtype in two case-control studies. Int J Cancer. 2013;132: 2640-2647.
47. Callahan CL, Hofmann JN, Corley DA, et al. Obesity and renal cell carcinoma risk by histologic subtype: A nested case-control study and meta-analysis. Cancer Epidemiol. 2018;56: 31-37.
48. Gansler T, Fedewa SA, Flanders WD, Pollack LA, Siegel DA, Jemal A. Prevalence of Cigarette Smoking among Patients with Different Histological Types of Kidney Cancer. Cancer Epidemiol Biomarkers Prev. 2020.
49. Colt JS, Hofmann JN, Schwartz K, et al. Antihypertensive medication use and risk of renal cell carcinoma. Cancer Causes Control. 2017;28: 289-297.
50. Le Marchand L, Wilkens LR. Design considerations for genomic association studies: importance of gene-environment interactions. Cancer Epidemiol Biomarkers Prev. 2008;17: 263-267.
51. Hunter DJ. Gene-environment interactions in human diseases. Nat Rev Genet. 2005;6: 287-298.
52. Kraft P, Hunter D. Integrating epidemiology and genetic association: the challenge of gene-environment interaction. Philos Trans R Soc Lond B Biol Sci. 2005;360: 1609-1616.
53. Antwi SO, Eckel-Passow JE, Diehl ND, et al. Alcohol consumption, variability in alcohol dehydrogenase genes and risk of renal cell carcinoma. Int J Cancer. 2018;142: 747-756.
54. Deckers IA, van den Brandt PA, van Engeland M, et al. Polymorphisms in genes of the renin-angiotensin-aldosterone system and renal cell cancer risk: interplay with hypertension and intakes of sodium, potassium and fluid. Int J Cancer. 2015;136: 1104-1116.
55. Karami S, Brennan P, Navratilova M, et al. Vitamin d pathway genes, diet, and risk of renal cell carcinoma. Int J Endocrinol. 2010;2010: 879362.
56. Smits KM, Schouten LJ, van Dijk BA, et al. Polymorphisms in genes related to activation or detoxification of carcinogens might interact with smoking to increase renal cancer risk: results from The Netherlands Cohort Study on diet and cancer. World J Urol. 2008;26: 103-110.
57. Melkonian SC, Daniel CR, Ye Y, et al. Gene-environment interaction of genome-wide association study-identified susceptibility loci and meat-cooking mutagens in the etiology of renal cell carcinoma. Cancer. 2016;122: 108-115.
58. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47: D941-d947.
59. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2: 401-404.
60. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6: pl1.
61. Hakimi AA, Chen YB, Wren J, et al. Clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma. Eur Urol. 2013;63: 848-854.
62. Hakimi AA, Mano R, Ciriello G, et al. Impact of recurrent copy number alterations and cancer gene mutations on the predictive accuracy of prognostic models in clear cell renal cell carcinoma.

J Urol. 2014;192: 24-29.
63. Hakimi AA, Ostrovnaya I, Reva B, et al. Adverse outcomes in clear cell renal cell carcinoma with mutations of 3p21 epigenetic regulators BAP1 and SETD2: a report by MSKCC and the KIRC TCGA research network. Clin Cancer Res. 2013;19: 3259-3267.
64. Manley BJ, Zabor EC, Casuscelli J, et al. Integration of Recurrent Somatic Mutations with Clinical Outcomes: A Pooled Analysis of 1049 Patients with Clear Cell Renal Cell Carcinoma. Eur Urol Focus. 2017;3: 421-427.
65. Patard JJ, Fergelot P, Karakiewicz PI, et al. Low CAIX expression and absence of VHL gene mutation are associated with tumor aggressiveness and poor survival of clear cell renal cell carcinoma. Int J Cancer. 2008;123: 395-400.
66. Voss MH, Reising A, Cheng Y, et al. Genomically annotated risk model for advanced renal-cell carcinoma: a retrospective cohort study. Lancet Oncol. 2018;19: 1688-1698.
67. Wang X, Steensma JT, Bailey MH, Feng Q, Padda H, Johnson KJ. Characteristics of The Cancer Genome Atlas cases relative to U.S. general population cancer cases. Br J Cancer. 2018;119: 885-892.
68. Spratt DE, Chan T, Waldron L, et al. Racial/Ethnic Disparities in Genomic Sequencing. JAMA Oncol. 2016;2: 1070-1074.
69. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn). 2015;19: A68-77.
70. van den Brandt PA, Goldbohm RA, van ‘t Veer P, Volovics A, Hermus RJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. J Clin Epidemiol. 1990;43: 285-295.
71. Hogervorst JG, Godschalk RW, van den Brandt PA, et al. DNA from nails for genetic analyses in large-scale epidemiologic studies. Cancer Epidemiol Biomarkers Prev. 2014;23: 2703-2712.
72. Goldbohm RA, Van den Brandt PA, Dorant E. Estimation of the coverage of Dutch municipalities by cancer registries and PALGA based on hospital discharge data. Tijdschr Soc Gezondheidsz. 1994;72: 80-84.
73. Eble J, Sauter G, Epstein J, Sesterhenn I. World Health Organization Classification of Tumours. Pathology and Genetics. Tumours of the Urinary System and Male Genital Organs. Lyon: IARC Press, 2004.

CHAPTER 2

Type 2 diabetes and its treatment and renal cell cancer risk

Jeroen AA van de Pol, Piet A van den Brandt and Leo J Schouten

Submitted for publication

CHAPTER 3

Etiologic heterogeneity of clear-cell and papillary renal cell carcinoma in the Netherlands Cohort Study

Jeroen AA van de Pol, Lisa George, Piet A van den Brandt, Marcella MLL Baldewijns and Leo J Schouten

Int. J. Cancer. 2020; 1-10

Abstract

At present, mostly case-control and retrospective studies have investigated the association between etiologic risk factors and the development of histologic subtypes of renal cell carcinoma (RCC). Therefore, we assessed the heterogeneity between body mass index (BMI), cigarette smoking, alcohol consumption and hypertension across clear-cell RCC (ccRCC) and papillary RCC (pRCC) risk in the prospective Netherlands Cohort Study on diet and cancer (NLCS). In 1986, 120,852 participants aged 55-69 completed a self-administered questionnaire on diet and other risk factors for cancer. Participants were followed-up for cancer through record linkage. Tumor histology was assessed through centralized revision by two experienced uropathologists. After 20.3 years of follow-up, 384 histologically verified RCC cases, including 315 ccRCC and 46 pRCC cases, and 4144 subcohort members were eligible for case-cohort analysis. Hazard ratios (HR) and 95% confidence intervals (CIs) were estimated by multivariable-adjusted proportional hazards models. Overall, BMI was associated positively with ccRCC risk, but inversely with pRCC risk. Cigarette smoking was associated with an increased ccRCC, but a decreased pRCC risk. Alcohol consumption was inversely associated with both ccRCC and pRCC risk. Hypertension was associated with an increased risk of both ccRCC and pRCC. Statistically significant etiologic heterogeneity was observed for BMI, BMI change since age 20, and smoking duration in current smokers across ccRCC and pRCC risk. In conclusion, we observed potential heterogeneity for BMI, BMI change and smoking duration across ccRCC and pRCC risk.

Background

Kidney cancer consists primarily of adenocarcinomas that arise in the renal parenchyma, commonly referred to as renal cell carcinomas (RCC) [1]. RCC is comprised of various entities defined by a distinct tumor histology, chromosomal alterations and molecular pathways [2]. The most common subtypes are clear cell (approximately 70% of all RCC) and papillary renal cell carcinoma (10-15%) [2].

Established modifiable risk factors for RCC include cigarette smoking, excess body weight and hypertension [3-8]. Furthermore, alcohol consumption has been associated with a decreased RCC risk in multiple prospective epidemiological studies [9, 10]. Even though histological subtypes of RCC have been formally recognized for more than two decades [11], data on etiologic risk factors linked to these subtypes remains sparse [1]. Previous studies have found evidence for a potential heterogeneity between the risk of ccRCC and pRCC for body mass index [12-15], and antihypertensive medication [16]. At present, no histological heterogeneity has been found in relation to cigarette smoking or hypertension status [12, 14]. Lastly, to our knowledge, no studies have directly assessed the etiologic risk heterogeneity between alcohol consumption and histologic RCC subtypes yet. The current available evidence on the heterogeneity between histologic RCC subtypes for these modifiable risk factors is solely based on information from (nested) case-control studies [12-14, 16] and three retrospective studies [15, 17, 18]. Additional evidence from large-scale prospective cohort studies may aid in uncovering potential heterogeneity between these established risk factors and ccRCC and pRCC development.

In the Netherlands Cohort Study (NLCS) on diet and cancer, a large nationwide prospective cohort study, we were able to assess tumor histology through a centralized revision by two experienced pathologists. With this information, we aimed to investigate the heterogeneity of associations between ccRCC and pRCC for the main established etiologic risk factors of RCC, namely BMI, cigarette smoking, alcohol consumption and hypertension.

Methods

Study population

The NLCS is a nation-wide prospective cohort study initiated in September 1986 with the inclusion of 58,279 men and 62,573 women between the ages of 55-69 years. The study design has been described in detail elsewhere [19]. In short, the study is a prospective cohort study initiated to investigate the association between diet and cancer risk. A case-cohort design was used for efficiency in follow-up for vital status and data processing. A subcohort of 5000 participants, of which 2411 men and 2589 women, was randomly sampled from the full cohort at baseline to estimate person-time at risk for the entire cohort. All participants were followed up by computerized record linkage with the Netherlands Cancer Registry (NCR), the Netherlands Pathology Registry (PALGA), and cause of death from Statistics Netherlands (CBS). In addition, participants were regularly followed up for migration and vital status. Follow-up for vital status of the subcohort was nearly 100% complete after 20.3 years and the completeness of cancer follow-up through record linkage is estimated to be over 96% [20]. The institutional review boards of Maastricht University (Maastricht) and the Netherlands Organization for Applied Scientific Research TNO (Zeist) approved the NLCS. The NLCS was conducted in accordance with the Declaration of Helsinki. By completing and returning the baseline questionnaire, participants agreed to participate in the NLCS.

In total, 608 RCC cases were identified within the NLCS between 1986 and 2006. Histologically confirmed RCC cases were eligible for the collection of formalin-fixed paraffin-embedded (FFPE) tumor tissues. Overall, FFPE tumor tissues were collected for 454 (79.8%) of the 568 eligible cases [21]. Tumor histology was revised by two experienced pathologists according to the WHO-classification of RCC tumors [2]. Of the 454 RCC cases, 366 (80.6%) were clear-cell (cc)RCC cases, 60 (13.2%) papillary (p)RCC cases, 15 (3.3%) chromophobe RCC cases, and 13 (2.9%) other or undefined RCC cases. Further classification of pRCC cases resulted in 35 (7.7%) type 1 pRCC, 24 (5.3%) type 2 pRCC, and 1 (0.2%) undefined pRCC. To maintain sufficient power in the analyses, type 1 and type 2 pRCC were combined into one category. Chromophobe RCC and other or undefined RCC cases were not assessed due to the insufficient number of cases.

Cohort members with prevalent cancer at baseline, except skin cancer, and incomplete or inconsistent information on exposure variables and a priori selected confounders were excluded from analyses. In total, 515 RCC cases (International Classification for Oncology 3: C64.9) and 4144 subcohort members were included in the analyses. Of the 384 included RCC cases with confirmed tumor histology 315 (82.0%) were ccRCC cases and 46 (12.0%) were pRCC cases.

Exposure assessment

All participants completed a mailed, self-administered questionnaire at baseline on dietary habits and other risk factors for cancer. By completing and returning the baseline questionnaire, individuals agreed to participate in the NLCS. From this questionnaire information was derived on anthropometric measures, smoking habits, dietary habits and medical conditions.

Baseline BMI (kg/m^2) was calculated using weight at baseline and the self-reported height squared. To calculate the BMI at age 20, the self-reported weight at age 20 was used in combination with the self-reported height at baseline. Change in BMI since age 20 was calculated by subtracting the BMI at age 20 from the BMI at baseline. In addition, men reported their trouser size and women reported their skirt size as a proxy for body composition [22]. Questions on cigarette, cigar and pipe smoking were used to assess smoking status, smoking quantity and smoking duration. Questions on beer, red wine, white wine, sherry, fortified wines, liqueur, and liquor were used to assess the consumption of alcohol. Participants who consumed alcoholic beverages less than once a month were considered non-users. Standard glass sizes were defined as 200 ml for beer, 105 ml for wine, 80 ml for sherry, and 45 ml for both liqueur and liquor [23]. These values corresponded to 8, 10, 11, 7 and 13 grams of alcohol, respectively. Mean daily alcohol consumption was calculated by multiplying the consumption frequency and the standardized item unit of each alcoholic beverage. Information from the questionnaire was also used to define stable abstainers and stable users of alcohol. Stable abstainers were defined as participants that reported no alcohol consumption 5 years before baseline. Stable users were defined as participants who reported that they drank equal amounts of alcoholic beverages 5 years before baseline. The diagnosis of hypertension was derived from a question on whether the participant was diagnosed with hypertension preceding baseline by a physician. In addition, participants were asked to report the use of any drugs for a period longer than 6 months. From this information, the use of antihypertensive medication was extracted.

Statistical analyses

Cox proportional hazard models were used to estimate sex-adjusted and multivariable-adjusted Hazard Ratio's (HRs) and 95% confidence intervals (CIs). Stata statistical software: release 15 (StataCorp., 2017, College Station, TX) was used for all analyses. Analyses were adjusted for smoking status (never/former/current), smoking duration (y, continuous, centered), smoking frequency (cig/d, continuous, centered), pipe and/or cigar smoking (never/former/current), alcohol consumption (g/d, continuous), body mass index (kg/m^2, continuous), diabetes status (no/yes) when applicable. Analyses on BMI change were additionally adjusted for BMI at age 20. As proposed by Leffondré et al. [24], smoking duration and smoking frequency were centered to avoid multicollinearity with smoking status. Analyses on smoking cessation were additionally adjusted for cigarette-years, calculated by multiplying smoking frequency with smoking duration, to resolve multicollinearity between smoking duration and smoking cessation. Fruit consumption, vegetable consumption and use of antihypertensive medication were included in models as potential confounders if they altered HRs for RCC risk by more than 10%. None of these potential confounders satisfied this condition, and were, therefore, not included in models as a confounding factor. The use of antihypertensive medication was studied as a potential risk factor based on findings from previous studies [5, 25].

Person-years at risk were calculated from baseline until registration of RCC or until date of censoring by death, emigration, loss to follow-up or end of follow-up, whichever occurred first. The proportional hazards assumption was tested with scaled Schoenfeld residuals and log-log curves. The proportional hazard assumption was violated for age, BMI and smoking frequency when using time-on-study as timescale. To resolve this issue, age-on-study was used as timescale with smoking frequency as a time-varying covariate. Standard errors were calculated using the robust Huber-White sandwich estimator, similar to the variance-covariance estimator by Barlow [26], to account for additional variance introduced by sampling a subcohort from the full cohort.

Test for heterogeneity of associations were performed to evaluate differences between ccRCC and pRCC risk for all etiologic risk factors using the competing risks procedure in Stata. *P*-values were calculated with a method developed for the case-cohort design based on bootstrapping. This procedure has been described in more detail elsewhere [27, 28]. All tests were performed two-sided and *P*-values <0.05 were considered statistically significant.

Results

The age at baseline of RCC cases was slightly lower, compared to the subcohort (Table 1). In addition, RCC cases were predominantly men, had a slightly increased mean BMI at baseline, BMI at age 20 and trouser and skirt size, were more often current and former smokers, consumed more alcohol and more often reported a diagnosis of hypertension and anti-hypertensive medication use, compared to the subcohort. Compared to ccRCC cases, papillary RCC cases, were more often men, had a lower BMI at baseline, had a lower trouser and skirt size, were more often former cigarette smokers and pipe or cigar smokers, had a higher cigarette smoking duration and frequency, and consumed less alcohol.

Overall, results between sex-adjusted and multivariable-adjusted cox-regression models did not indicate large differences. Therefore, sex-adjusted analyses are presented in the

supplements (Supplementary Tables 1.1 – 1.4).

Table 1 - Baseline characteristics of the subcohort and Renal Cell Carcinoma cases in the Netherlands Cohort Study on diet and cancer, 1986-2006

	Subcohort	Renal Cell Carcinoma		
Baseline characteristics:		Overall	ccRCC	pRCC
Total, (n)	4144	515	315	46
Age at baseline, (years)	61.3 (4.2)	60.9 (3.9)	60.7 (3.9)	61.1 (3.9)
Male sex, (n, %)	2039 (49.2)	337 (65.4)	200 (63.5)	40 (87.0)
Body Mass Index at baseline, (kg/m2)	25.0 (3.1)	25.4 (3.0)	25.5 (3.0)	24.6 (2.2)
Body Mass Index at age 20, (kg/m2)	21.5 (2.6)	21.7 (2.6)	21.9 (2.7)	21.7 (2.1)
Trouser size in men at baseline, (size)	51.5 (4.3)	52.1 (3.1)	52.2 (3.4)	51.6 (2.7)
Skirt size in women at baseline, (size)	43.5 (3.0)	44.3 (3.0)	44.2 (2.7)	42.8 (1.8)
Cigarette smoking status, (n, %)				
Never smokers	1524 (36.8)	136 (26.4)	85 (27.0)	10 (21.7)
Former smokers	1466 (35.4)	219 (42.5)	136 (43.2)	22 (47.8)
Current smokers	1154 (27.9)	160 (31.1)	94 (29.8)	14 (30.4)
Ever cigarette smokers only				
Smoking duration, (years)	31.8 (12.2)	32.1 (12.0)	31.7 (11.9)	36.5 (10.6)
Smoking frequency, (cig/d)	15.4 (10.3)	17.1 (12.2)	16.5 (11.5)	18.0 (12.0)
Pipe and/or cigar smoking				
Never pipe or cigar smoker, (n, %)	3559 (85.9)	414 (80.4)	248 (78.7)	32 (69.6)
Former pipe or cigar smoker, (n, %)	308 (7.4)	66 (12.8)	43 (13.7)	10 (21.7)
Current pipe or cigar smoker, (n, %)	277 (6.7)	35 (6.8)	24 (7.6)	4 (8.7)
Alcohol intake, (g/d)[a]	13.5 (15.1)	15.2 (15.3)	15.0 (15.0)	13.8 (12.3)
Diagnosis of hypertension, (n, %)	1093 (26.4)	161 (31.3)	99 (31.4)	16 (34.8)
Use of antihypertensive medication, (n, %)	856 (20.7)	124 (24.1)	80 (25.4)	11 (23.9)

Abbreviations: ccRCC: clear cell Renal Cell Carcinoma, pRCC: papillary Renal Cell Carcinoma

[a] In consumers only.

BMI

In multivariable-adjusted analyses, a positive association was observed between BMI and RCC risk (Table 2). Furthermore, we observed an U-shaped association between BMI at age 20 and RCC risk. BMI change per one kg/m^2 increment since age 20 was associated with a non-statistically significantly increased RCC risk. Trouser and skirt size were associated with an increased risk of RCC across increasing size categories (p-trend: 0.02, 0.005 for trouser and skirt size, respectively).

In general, an increasing ccRCC risk was observed across increasing BMI categories. The strength of the associations was slightly elevated compared to associations observed for overall RCC risk. In addition, a statistically significantly increased risk was found per one kg/m^2 increase (HR 1.04, 95%CI 1.01-1.08). For pRCC risk, associations became increasingly inverse across increasing BMI categories. A borderline significant inverse association was found for pRCC per kg/m^2 increase (HR 0.91, 95%CI 0.82-1.00). We observed statistically significant heterogeneity across ccRCC and pRCC for baseline BMI (per kg/m^2; $p_{heterogeneity}$: 0.02). Furthermore, an U-shaped association was found between BMI at age 20 and ccRCC,

while no clear association was found for pRCC. Furthermore, an increase in change in BMI since age 20 was associated with an increase in risk for ccRCC, and a decrease in pRCC risk. These differences in BMI change since 20 were statistically significant in tests for heterogeneity ($p_{heterogeneity}$: 0.03). Trouser size in men was associated with a statistically significantly increased ccRCC risk per size increase, while no association was found with pRCC risk. However, no heterogeneity of associations was observed ($p_{heterogeneity}$: 0.18). No analyses were performed for skirt size in pRCC, because of the limited number of female pRCC cases (n=5).

Table 2 - Multivariable-adjusted Cox proportional hazard models with age as timescale for the association between Body Mass Index (BMI) and risk of clear-cell Renal Cell Carcinoma and papillary Renal Cell Carcinoma in the Netherlands Cohort Study on diet and cancer, 1986-2006

Char.		clear-cell Renal Cell Carcinoma			papillary Renal Cell Carcinoma			
	Sub-cohort	Cases	Multivariable-adjusted[a]		Cases	Multivariable-adjusted[a]		P for
	person-years	No.	HR	95% CI	No.	HR	95% CI	hetero-geneity[b]
BMI at baseline (kg/m^2)								
<23	16723	55	1	Ref.	11	1	Ref.	
23-<25	21298	90	1.14	(0.81-1.63)	18	0.92	(0.42-2.04)	
25-<27	16372	81	1.30	(0.90-1.86)	10	0.60	(0.24-1.48)	
≥27	15477	89	1.61	(1.13-2.30)	7	0.50	(0.19-1.33)	0.62 [f]
p trend			0.005			0.09		
Cont. (per kg/m^2)	69871	315	1.04	(1.01-1.08)	46	0.91	(0.82-1.00)	0.02
BMI at age 20 (kg/m^2) [c]								
<20.0	15384	61	1.39	(0.93-2.09)	6	0.75	(0.26-2.17)	
20.0-<21.5	14576	45	1	Ref.	9	1	Ref.	
21.5-<23	14558	67	1.50	(1.01-2.23)	10	1.08	(0.42-2.83)	
≥23	15549	85	1.76	(1.21-2.57)	8	0.77	(0.29-2.04)	0.56
p trend			0.05			0.98		
Cont. (per kg/m^2)	60067	258	1.05	(1.00-1.10)	33	1.01	(0.90-1.13)	0.70
BMI change since age 20 (kg/m^2) [d]								
<1.5	14892	60	1	Ref.	8	1	Ref.	
1.5-<3.5	16693	73	1.08	(0.75-1.57)	16	1.31	(0.50-3.43)	
3.5-<5.5	13659	58	1.15	(0.77-1.72)	7	0.69	(0.22-2.14)	
≥5.5	14823	67	1.34	(0.89-2.02)	2	0.16	(0.03-0.82)	0.80
p trend			0.15			0.005		
Cont. (per kg/m^2)	60067	258	1.03	(0.99-1.08)	33	0.89	(0.75-0.98)	0.03

Continues on next page

Table 2 - Continued

		clear-cell Renal Cell Carcinoma			papillary Renal Cell Carcinoma			
	Sub-cohort	Cases	Multivariable-adjusted[a]		Cases	Multivariable-adjusted[a]		P for
	person-years	No.	HR	95% CI	No.	HR	95% CI	hetero-geneity[b]
Trouser size - Men								
<50	4745	26	1.16	(0.68-1.99)	8	1.86	(0.62-5.59)	
50-<52	6736	32	1	Ref.	6	1	Ref.	
52-<54	10137	63	1.31	(0.84-2.04)	12	1.13	(0.42-3.07)	
54-<56	5553	41	1.54	(0.95-2.51)	8	1.46	(0.51-4.17)	
≥56	2756	24	1.90	(1.09-3.33)	3	0.98	(0.23-4.11)	0.94
p trend			0.03			0.57		
Cont. (per size)	29928	186	1.06	(1.01-1.12)	37	1.00	(0.95-1.05)	0.18
Skirt size - women [e]								
<42	7012	15	0.87	(0.45-1.69)	1	-		
42-<44	9372	24	1	Ref.	1	1		
44-<46	10272	31	1.16	(0.67-2.01)	3	-		
46-<48	6529	27	1.52	(0.87-2.67)	0	-		
≥48	3774	17	1.55	(0.81-2.98)	0	-		-
p trend			0.04			-		
Cont. (per size)	36958	114	1.07	(1.01-1.13)	5	-		-

[a] Additionally adjusted for: smoking status (never/former/current), smoking duration (continuous, centered) smoking frequency (continuous, centered), pipe and/or cigar smoking (never/former/current), hypertension status (no/yes), alcohol consumption (g/d, continuous), diabetes (no/yes), models included a time-varying covariate for smoking frequency because of a potential violation of the proportional hazards assumption

[b] Based on multivariable-adjusted models

[c] Analysis with BMI at age 20 have a restricted number of cases due to missing values

[d] Additionally adjusted for BMI at age 20

[e] Not adjusted for pipe and/or cigar smoking due to unstable estimates for these confounding factors

[f] Models failed to converge more than 10 times during heterogeneity analyses (1000 replications). Models and intrinsic standard errors are based solely on successful bootstraps.

Cigarette smoking

In multivariable-adjusted models, cigarette smoking status was associated with an increased risk of RCC (Table 3). Associations persisted after adjustment for smoking frequency and duration. Restricting analyses to exclusively cigarette smokers strengthened associations in current cigarette smokers. No clear association was found with smoking duration and smoking frequency. Smoking cessation was associated with a decreased RCC risk in categorized analyses.

Similar to RCC overall, an increased ccRCC risk was found in both former (HR 1.26, 95%CI 0.91-1.74) and current smokers (HR 1.41, 95%CI 1.01-1.97). Associations remained similar after adjusting for smoking frequency and duration, and when restricting analyses

to exclusively cigarette smokers. No clear association was found between smoking status and pRCC risk. However, when restricting analyses to exclusively cigarette smokers, an inverse association with pRCC risk was found in former (HR 0.66; 95%CI 0.22-1.98) and current smokers (HR 0.46, 95%CI 0.11-1.90). Heterogeneity tests between ccRCC and pRCC risk were not able to provide reliable estimations for analyses on smoking status. No clear association was found between smoking frequency and smoking duration and ccRCC risk. An increased pRCC risk was observed in cases with increasing smoking frequency and duration, although not statistically significant. Statistically significant heterogeneity was observed for smoking duration in current smokers per 5-year increment ($p_{heterogeneity}$: 0.04), but not in former smokers ($p_{heterogeneity}$: 0.17). While there were indications for a decreased ccRCC and pRCC risk with increasing categories of duration of smoking cessation, solely analyses on pRCC risk suggested a potential inverse association per 5 years increase. Tests for heterogeneity did not show statistically significant differences across ccRCC and pRCC for smoking cessation ($p_{heterogeneity}$: 0.98).

Table 3 - Multivariable-adjusted Cox proportional hazard models with age as timescale for the association between cigarette smoking and risk of clear-cell Renal Cell Carcinoma and papillary Renal Cell Carcinoma in the Netherlands Cohort Study on diet and cancer, 1986-2006

		clear-cell Renal Cell Carcinoma			papillary Renal Cell Carcinoma			
	Sub-cohort	Cases	Multivariable-adjusted[a]		Cases	Multivariable-adjusted[a]		P for
	person-years	No.	HR	95% CI	No.	HR	95% CI	hetero-geneity[b]
Unadjusted for cigarette smoking frequency and duration								
Cigarette smoking status								
Never	27409	85	1	Ref.	10	1	Ref.	
Former	24499	136	1.26	(0.91-1.74)	22	0.84	(0.35-2.02)	
Current	17963	94	1.41	(1.01-1.97)	14	1.08	(0.47-2.51)	0.73[g]
p trend			0.05			0.78		
Never	27409	85	1	Ref.	10	1	Ref.	
Former	24499	136	1.26	(0.90-1.78)	22	0.97	(0.39-2.46)	
Current	17963	94	1.39	(0.98-1.97)	14	0.65	(0.24-1.77)	0.75[g]
p trend			0.06			0.39		
Adjusted for cigarette smoking frequency and duration [c]								
Cigarette smoking status								
Never	26090	73	1	Ref.	9	1	Ref.	
Former	19102	93	1.37	(0.95-1.97)	10	0.66	(0.22-1.98)	
Current	15348	82	1.54	(1.07-2.22)	13	0.46	(0.11-1.90)	0.71[g]
p trend			0.02			0.28		

Continues on next page

Table 3 - Continued

		clear-cell Renal Cell Carcinoma			papillary Renal Cell Carcinoma			
	Sub-cohort	Cases	Multivariable-adjusted[a]		Cases	Multivariable-adjusted[a]		P for
	person-years	No.	HR	95% CI	No.	HR	95% CI	hetero-geneity[b]
Solely in exclusive cigarette smokers								
Smoking frequency [d]								
>0-<20 cig/d	21689	102	1	Ref.	10	1	Ref.	
≥20 cig/d	12761	73	1.01	(0.72-1.42)	13	1.67	(0.68-4.10)	0.18
Cont. (per 5 cig/d), former smoker	19102	93	1.03	(0.92-1.16)	10	1.10	(0.83-1.47)	0.42
Cont. (per 5 cig/d), current smoker	15348	82	1.03	(0.93-1.14)	13	1.22	(0.91-1.64)	0.27
Smoking duration [e]								
>0-<30 yrs	13701	65	1	Ref.	2	1	Ref.	
≥30 yrs	20749	110	0.90	(0.62-1.32)	21	4.96	(1.19-20.7)	-[h]
Cont. (per 5 yrs), former smoker	19102	93	1.03	(0.92-1.14)	10	1.67	(1.04-2.67)	0.17
Cont. (per 5 yrs), current smoker	15348	82	0.96	(0.85-1.08)	13	1.26	(0.63-2.51)	0.04
Number of years of smoking cessation [f]								
Current smoker	15348	82	1	Ref.	13	1	Ref.	
>0-<15	10444	53	0.91	(0.65-1.29)	7	0.76	(0.30-1.94)	
≥15	8657	40	0.88	(0.59-1.33)	3	0.44	(0.12-1.64)	
Never smoker	26090	73	0.65	(0.46-0.92)	9	1.62	(0.52-5.06)	0.95
p trend, excl. never smokers			0.49			0.22		
Cont. (per 5 yrs), former smokers	19102	93	0.98	(0.86-1.11)	10	0.87	(0.58-1.31)	0.98

[a] Additionally adjusted for: pipe and/or cigar smoking (never/former/current), Body Mass Index (continuous), hypertension status (no/yes), alcohol consumption (g/d, continuous), diabetes (no/yes), models included a time-varying covariate for smoking frequency because of a potential violation of the proportional hazards assumption

[b] Based on multivariable-adjusted models

[c] Additionally adjusted for cigarette smoking duration (years, centered) and cigarette smoking frequency (cig/d, centered)

[d] Additionally adjusted for smoking duration (years, centered)

[c] Additionally adjusted for smoking frequency (cig/d, centered)
[f] Additionally adjusted for cigarette-years
[g] Models failed to converge more than 10 times during heterogeneity analyses (1000 replications). Models and intrinsic standard errors are based solely on successful bootstraps.
[h] Estimates were unstable due to limited sample sizes in subcategories.

Alcohol

In multivariable-analyses, alcohol consumption was associated with a seemingly non-linear decreased risk of RCC, although not statistically significant (Table 4). In stable users and abstainers, associations were mostly inverse. The inverse association with RCC risk was the strongest in the category 5-<15 g/d, when compared to abstainers, in both analyses with all alcohol users and analyses with stable users and abstainers. HRs attenuated in categories above 15 g/d.

Associations were slightly stronger in analyses on ccRCC risk, compared to associations found for RCC overall. A non-linear association with ccRCC risk was found for alcohol consumption in stable alcohol users and abstainers. In analyses on pRCC risk, a seemingly non-linear association was observed with an increased pRCC risk at alcohol consumptions <15 g/d and a decreased pRCC risk at higher alcohol consumptions, when compared to abstainers. In stable alcohol users and abstainers, an inverse association was observed between alcohol consumption and pRCC risk. However, these analyses were performed on a very limited number of participants. Tests for heterogeneity of associations between ccRCC and pRCC were not able to provide reliable estimations for categorical analyses on alcohol consumption. No heterogeneity was observed between ccRCC and pRCC risk in continuous analyses ($p_{heterogeneity}$ range: 0.85-0.86).

Hypertension

In multivariable analyses, self-reported hypertension and the self-reported use of antihypertensive medication were associated with an increased RCC risk (table 5). Risk estimates were elevated in participants who reported both hypertension and the use of antihypertensive medication.

Similar to analyses on RCC overall, hypertension and the use of anti-hypertensive medication risks were consistently associated with an increased risk for both ccRCC and pRCC. The observed associations were the strongest with pRCC. Participants who reported both hypertension and the use of antihypertensive medication had a strongly elevated risk of both ccRCC and pRCC. Overall, tests for heterogeneity indicated no differences between ccRCC and pRCC risk regarding the self-reported hypertension and use of anti-hypertensive medication ($p_{heterogeneity}$ range: 0.62-0.82).

Table 4 - Multivariable-adjusted Cox proportional hazard models with age as timescale for the association between alcohol and risk of clear-cell Renal Cell Carcinoma and papillary Renal Cell Carcinoma in the Netherlands Cohort Study on diet and cancer, 1986-2006

		clear-cell Renal Cell Carcinoma			papillary Renal Cell Carcinoma			
	Sub-cohort	Cases	Multivariable-adjusted[a]		Cases	Multivariable-adjusted[a]		P for
	person-years	No.	HR	95% CI	No.	HR	95% CI	hetero-geneity[b]
Alcohol consumption (g/d)								
Abstainer	16613	76	1	Ref.	7	1	Ref.	
>0-<5	20731	79	0.77	(0.56-1.07)	13	1.42	(0.55-3.66)	
≥5-<15	15589	64	0.71	(0.50-1.00)	14	1.48	(0.57-3.84)	
≥15-<30	10810	63	0.86	(0.59-1.25)	8	0.89	(0.32-2.46)	
≥30	6127	33	0.77	(0.49-1.23)	4	0.67	(0.19-2.36)	0.62 [d]
p trend			0.37			0.32		
Cont. (per 5 g/d)	69871	315	0.98	(0.94-1.03)	46	0.94	(0.86-1.04)	0.85
Alcohol consumption among stable abstainers/users (g/d) [c]								
Abstainer	12986	57	1	Ref.	7	1	Ref.	
>0-<5	11867	46	0.86	(0.57-1.28)	6	0.81	(0.27-2.42)	
≥5-<15	9435	42	0.87	(0.57-1.34)	6	0.86	(0.27-2.73)	
≥15-<30	5832	38	1.09	(0.67-1.77)	3	0.58	(0.15-2.30)	
≥30	3287	20	1.00	(0.55-1.83)	2	0.75	(0.14-3.93)	- [d]
p trend			0.82			0.57		
Cont. (per 5 g/d)	43407	203	1.00	(0.94-1.05)	24	0.99	(0.85-1.15)	0.86

[a] Additionally adjusted for: smoking status (never/former/current), smoking duration (continuous, centered), smoking frequency (continuous, centered), pipe and/or cigar smoking (never/former/current), hypertension status (no/yes), body mass index (kg/m2, continuous), diabetes (no/yes), models included a time-varying covariate for smoking frequency because of a potential violation of the proportional hazards assumption

[b] Based on multivariable-adjusted models

[c] Stable abstainers were defined as participants that reported no alcohol consumption 5 years before baseline. Stable users were defined as participants who reported that they drank equal amounts of beer or other alcoholic beverages 5 years before baseline.

[d] Models failed to converge more than 10 times during heterogeneity analyses (1000 replications). Models and intrinsic standard errors are based solely on successfull bootstraps.

Table 5 - Multivariable-adjusted Cox proportional hazard models with age as timescale for the association between hypertension and risk of clear-cell Renal Cell Carcinoma and papillary Renal Cell Carcinoma in the Netherlands Cohort Study on diet and cancer, 1986-2006

		clear-cell Renal Cell Carcinoma			papillary Renal Cell Carcinoma			
	Sub-cohort	Cases	Multivariable-adjusted[a]		Cases	Multivariable-adjusted[a]		P for
	person-years	No.	HR	95% CI	No.	HR	95% CI	hetero-geneity[b]
Self-reported hypertension diagnosis								
No	51868	216	1	Ref.	30	1	Ref.	
Yes	18002	99	1.32	(1.02-1.69)	16	1.95	(1.03-3.67)	0.67
Use of antihypertensive medication								
No	56463	235	1	Ref.	35	1	Ref.	
Yes	13408	80	1.45	(1.10-1.91)	11	1.47	(0.70-3.08)	0.82
Hypertension status & use of antihypertensive medication								
No hyp or no med	59848	251	1	Ref.	35	1	Ref.	
Hyp and med	10023	64	1.56	(1.16-2.09)	11	2.41	(1.15-5.01)	0.62

Abbreviations: hyp: Hypertension status, med: use of medication.

[a] Additionally adjusted for: smoking status (never/former/current), smoking duration (continuous, centered), smoking frequency (continuous, centered), pipe and/or cigar smoking (never/former/current), alcohol consumption (g/d, continuous), body mass index (kg/m2, continuous), diabetes (no/yes), models included a time-varying covariate for smoking frequency because of a potential violation of the proportional hazards assumption

[b] Based on multivariable-adjusted models

Discussion

In this large-scale prospective cohort study we investigated the etiologic heterogeneity between BMI, smoking, alcohol consumption and hypertension across ccRCC and pRCC risk. We observed statistically significant heterogeneity of associations across ccRCC and pRCC for BMI, BMI change since age 20 and smoking duration in current smokers. We observed no heterogeneity across histologic subtypes for alcohol consumption and hypertension.

There is a growing body of evidence regarding a potential subtype-specific association between BMI and RCC risk. In multiple studies, an association with obesity was observed with ccRCC, and not with pRCC risk [12, 14, 29]. In contrast, one study found no difference in BMI across the development of ccRCC and pRCC [15]. Other studies, which only reported differences between the occurrence of clear-cell RCC versus other histologic RCC subtypes combined, also reported potential histologic differences related to obesity [13, 30, 31]. In our study, we found evidence for heterogeneity across ccRCC and pRCC risk in continuous analyses for BMI at baseline, as well as for BMI change since age 20. Even though we did not observe statistically significant heterogeneity in categorical analyses, large differences in estimates were observed for BMI categories across ccRCC and pRCC risk. We report a similar association, compared to previous studies, between BMI and ccRCC risk. For pRCC risk, however, we observed consistent inverse associations regarding BMI, while previous studies found no association [12, 14, 29]. This may due to the limited power of our pRCC analyses, as we report results with wide confidence intervals for pRCC. The consistent report of heterogeneity across multiple studies, does provide an indication of etiologic differences across RCC subtypes regarding BMI.

Several plausible mechanisms may explain the observed association between BMI and ccRCC risk specifically. Clear-cell RCC commonly harbors somatic mutations in the *von Hippel-Lindau (VHL)* tumor-suppressor gene [32]. Inactivation of *VHL* leads to upregulation of the type 1 insulin-like growth factor receptor (IGF1R) in RCC cells [33]. Obesity has been related to hormonal changes in the body, including an increase in circulating levels of insulin-like growth factor-1 (IGF-1) [34]. IGF-1 strongly stimulates cell proliferation, inhibits apoptosis, and can enhance angiogenesis [35]. Therefore, the inactivation of *VHL* in ccRCC and the obesity-related increase in IGF-1 may amplify the process of tumorigenesis. Another hypothesized mechanism may be the relationship between hypoxia and the development of ccRCC [36]. Physiologically, the kidney is sensitive to perturbations in oxygen levels due to the activation of the *VHL-HIF1A* pathway [37]. Somatic mutations in this pathway are known to lead to carcinogenesis due to defects in the hypoxia sensing mechanism, which is characteristic of ccRCC development [37]. Therefore, it is hypothesized that chronic hypoxia may exert similar effects due to the repeated activation of the hypoxia sensing mechanisms [36]. Obesity may achieve this due to the link to obstructive sleep apnea, which causes a state of hypoxia during sleep [38]. Therefore, the presence of hypoxia due to obstructive sleep apnea could, in part, explain the observed relationship between obesity and ccRCC in particular [39]. However, verification of the association between obstructive sleep apnea and ccRCC risk should still be addressed in future studies.

To our knowledge, three studies have investigated the potential for heterogeneity between cigarette smoking and the risk of histologic subtypes of RCC [12, 14, 17]. In these studies, no

heterogeneity of associations was found across ccRCC and pRCC for tobacco smoking status [12, 14, 17]. In our study, a positive association was observed between cigarette smoking status and ccRCC risk, but no clear association was found with pRCC risk. Patel *et al.* [17] investigated the heterogeneity between smoking frequency, duration and pack-years across ccRC C and pRCC in detail and found no evidence for heterogeneity. Similarly, we found no heterogeneity of associations for cigarette smoking frequency and smoking cessation. We observed a statistically significant heterogeneity for smoking duration in current smokers ($p_{heterogeneity}$: 0.04). However, this may be a chance finding as we observed a very skewed distribution of cigarette-smoking patterns in pRCC cases, which may have affected the heterogeneity estimates between ccRCC and pRCC.

The current evidence suggests that alcohol consumption is inversely associated with the risk of RCC [9, 10, 40, 41]. As of yet, the potential for histologic heterogeneity regarding alcohol consumption has remained unexplored. In our study, we observed inverse associations between alcohol consumption and both ccRCC and pRCC risk. No heterogeneity was found across ccRCC and pRCC in either overall alcohol consumers or stable users and abstainers. However, these findings need to be validated in future studies.

Previous research has indicated that patients with hypertension more often present with non-clear cell histology [42]. Even though slightly stronger associations were found between hypertension and pRCC, when compared to ccRCC, we observed no statistically significant heterogeneity of associations, which is in line with findings from Purdue *et al.* [12]. A previous study by Colt et al. [16] has indicated potential drug- and histology-specific associations between hypertension and RCC. In their study pRCC, but not ccRCC, was associated with long-term use of diuretics and calcium channel blockers [16]. In our study, we observed similar point estimates between antihypertensive medication use and ccRCC risk and pRCC risk. Unfortunately, due to the limited number of pRCC cases, we were unable to assess the risk related to different types of antihypertensive medication.

The present study had several strengths, including the prospective design, the detailed assessment of exposures and confounders at baseline, and the long duration of cancer follow-up. Our results for overall RCC, as detailed in the supplementary materials (Supplementary Tables 2.1-2.4), were in line with evidence from large-scale meta-analyses on obesity [43], smoking [3, 8], alcohol consumption [9, 41] and hypertension [7], which strengthens the credibility of our results. Furthermore, the differentiation between histological subtypes was based on the centralized revision by two experienced uropathologists, which improved the accuracy of the information on histologic RCC subtypes. However, our study also was subject to some limitations. Firstly, information on anthropometry and exposure status was self-reported by the participants at one single time point. Consequentially, there may have been measurement error due to self-report. Moreover, changes in exposure status during follow-up were not recorded. Secondly, we had a limited number of cases with a tumor histology other than ccRCC. As a result, we were unable to assess the etiologic heterogeneity in more detail for exposures (e.g. anti-hypertensive medication subtypes), histologic subtypes (e.g. chromophobe RCC) and to further categorize the pRCC subtypes (e.g. type 1 and type 2 pRCC). In particular, the inability to further stratify pRCC may have influenced results as these were classified as pRCC overall, while type 1 and type 2 pRCC subtypes are known to

possess molecular and clinicopathologically distinct characteristics [2, 44]. It may be possible that type 1 and type 2 pRCC are distinctly associated to the risk factors included in our analyses. Lastly, due to the low number of pRCC cases, we were not able to obtain reliable estimates in all heterogeneity tests.

In conclusion, the results of our study are suggestive of the presence of an etiologic heterogeneity regarding RCC subtypes. In particular, we observed that the association between BMI and ccRCC and pRCC differs. These results highlight the need for more detailed subtype-specific analyses when investigating risk factors for RCC. Evidence from studies on specific tumor histologies may help uncover mechanisms that play a role in the process of tumorigenesis for RCC by uncovering etiologic similarities and differences between tumor subtypes.

References

1. Chow WH, Dong LM, Devesa SS. Epidemiology and risk factors for kidney cancer. Nature reviews. Urology. 2010;7: 245-257.
2. Eble JN, Sauter G, Epstein JI, Sesterhenn IA. Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs. Lyon: IARC *Press*, 2004.
3. Hunt JD, van der Hel OL, McMillan GP, Boffetta P, Brennan P. Renal cell carcinoma in relation to cigarette smoking: meta-analysis of 24 studies. International journal of cancer. 2005;114: 101-108.
4. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. Lancet (London, England). 2008;371: 569-578.
5. Weikert S, Boeing H, Pischon T, et al. Blood pressure and risk of renal cell carcinoma in the European prospective investigation into cancer and nutrition. American journal of epidemiology. 2008;167: 438-446.
6. Chow WH, Gridley G, Fraumeni JF, Jr., Jarvholm B. Obesity, hypertension, and the risk of kidney cancer in men. The New England journal of medicine. 2000;343: 1305-1311.
7. Hidayat K, Du X, Zou SY, Shi BM. Blood pressure and kidney cancer risk: meta-analysis of prospective studies. Journal of hypertension. 2017;35: 1333-1344.
8. Cumberbatch MG, Rota M, Catto JW, La Vecchia C. The Role of Tobacco Smoke in Bladder and Kidney Carcinogenesis: A Comparison of Exposures and Meta-analysis of Incidence and Mortality Risks. European urology. 2016;70: 458-466.
9. Bellocco R, Pasquali E, Rota M, et al. Alcohol drinking and risk of renal cell carcinoma: results of a meta-analysis. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO. 2012;23: 2235-2244.
10. Lee JE, Hunter DJ, Spiegelman D, et al. Alcohol intake and renal cell cancer in a pooled analysis of 12 prospective studies. Journal of the National Cancer Institute. 2007;99: 801-810.
11. Bostwick DG, Eble JN. Diagnosis and classification of renal cell carcinoma. Urol Clin North Am. 1999;26: 627-635.
12. Purdue MP, Moore LE, Merino MJ, et al. An investigation of risk factors for renal cell carcinoma by histologic subtype in two case-control studies. International journal of cancer. 2013;132: 2640-2647.
13. Dal Maso L, Zucchetto A, Tavani A, et al. Renal cell cancer and body size at different ages: an Italian multicenter case-control study. American journal of epidemiology. 2007;166: 582-591.
14. Callahan CL, Hofmann JN, Corley DA, et al. Obesity and renal cell carcinoma risk by histologic subtype: A nested case-control study and meta-analysis. Cancer epidemiology. 2018;56: 31-37.
15. Lee WK, Lee SE, Hong SK, et al. Characteristics and prognostic value of papillary histologic subtype in nonmetastatic renal cell carcinoma in Korea: a multicenter study. Urology journal. 2014;11: 1884-1890.
16. Colt JS, Hofmann JN, Schwartz K, et al. Antihypertensive medication use and risk of renal cell carcinoma. Cancer causes & control : CCC. 2017;28: 289-297.
17. Patel NH, Attwood KM, Hanzly M, et al. Comparative Analysis of Smoking as a Risk Factor among Renal Cell Carcinoma Histological Subtypes. The Journal of urology. 2015;194: 640-646.
18. Lowrance WT, Thompson RH, Yee DS, Kaag M, Donat SM, Russo P. Obesity is associated with a higher risk of clear-cell renal cell carcinoma than with other histologies. BJU international.

2010;105: 16-20.
19. van den Brandt PA, Goldbohm RA, van ‘t Veer P, Volovics A, Hermus RJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. Journal of clinical epidemiology. 1990;43: 285-295.
20. Goldbohm RA, van den Brandt PA, Dorant E. Estimation of the coverage of Dutch municipalities by cancer registries and PALGA based on hospital discharge data1994.
21. Deckers IA, van Engeland M, van den Brandt PA, et al. Promoter CpG island methylation in ion transport mechanisms and associated dietary intakes jointly influence the risk of clear-cell renal cell cancer. Int J Epidemiol. 2017;46: 622-631.
22. Hughes LAE, Schouten LJ, Goldbohm RA, van den Brandt PA, Weijenberg MP. Self-reported clothing size as a proxy measure for body size. Epidemiology. 2009;20: 673-676.
23. Schouten LJ, van Dijk BA, Oosterwijk E, et al. Alcohol consumption and mutations or promoter hypermethylation of the von Hippel-Lindau gene in renal cell carcinoma. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2008;17: 3543-3550.
24. Leffondré K, Abrahamowicz M, Siemiatycki J, Rachet B. Modeling smoking history: a comparison of different approaches. American journal of epidemiology. 2002;156: 813-823.
25. Sanfilippo KM, McTigue KM, Fidler CJ, et al. Hypertension and obesity and the risk of kidney cancer in 2 large cohorts of US men and women. Hypertension (Dallas, Tex. : 1979). 2014;63: 934-941.
26. Barlow WE. Robust variance estimation for the case-cohort design. Biometrics. 1994;50: 1064-1072.
27. Wacholder S, Gail MH, Pee D, Brookmeyer R. Alternative Variance and Efficiency Calculations for the Case-Cohort Design. Biometrika. 1989;76: 117-123.
28. de Vogel S, Bongaerts BW, Wouters KA, et al. Associations of dietary methyl donor intake with MLH1 promoter hypermethylation and related molecular phenotypes in sporadic colorectal cancer. Carcinogenesis. 2008;29: 1765-1773.
29. Lipworth L, Morgans AK, Edwards TL, et al. Renal cell cancer histological subtype distribution differs by race and sex. BJU international. 2016;117: 260-265.
30. Donat SM, Salzhauer EW, Mitra N, Yanke BV, Snyder ME, Russo P. Impact of body mass index on survival of patients with surgically treated renal cell carcinoma. The Journal of urology. 2006;175: 46-52.
31. Steffens S, Ringe KI, Schroeer K, et al. Does overweight influence the prognosis of renal cell carcinoma? Results of a multicenter study. International journal of urology : official journal of the Japanese Urological Association. 2013;20: 585-592.
32. Gnarra JR, Tory K, Weng Y, et al. Mutations of the VHL tumour suppressor gene in renal carcinoma. Nature genetics. 1994;7: 85-90.
33. Yuen JSP, Cockman ME, Sullivan M, et al. The VHL tumor suppressor inhibits expression of the IGF1R and its loss induces IGF1R upregulation in human clear cell renal carcinoma. Oncogene. 2007;26: 6499-6508.
34. Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. Nat Rev Cancer. 2004;4: 579-591.
35. Khandwala HM, McCutcheon IE, Flyvbjerg A, Friend KE. The effects of insulin-like growth factors on tumorigenesis and neoplastic growth. Endocr Rev. 2000;21: 215-244.
36. Sharifi N, Farrar WL. Perturbations in hypoxia detection: a shared link between hereditary and sporadic tumor formation? Medical hypotheses. 2006;66: 732-735.

37. Schodel J, Grampp S, Maher ER, et al. Hypoxia, Hypoxia-inducible Transcription Factors, and Renal Cancer. European urology. 2016;69: 646-657.
38. Gibson GJ. Obstructive sleep apnoea syndrome: underestimated and undertreated. Br Med Bull. 2005;72: 49-65.
39. Gozal D, Ham SA, Mokhlesi B. Sleep Apnea and Cancer: Analysis of a Nationwide Population Sample. Sleep. 2016;39: 1493-1500.
40. Song DY, Song S, Song Y, Lee JE. Alcohol intake and renal cell cancer risk: a meta-analysis. Br J Cancer. 2012;106: 1881-1890.
41. Xu X, Zhu Y, Zheng X, Xie L. Does beer, wine or liquor consumption correlate with the risk of renal cell carcinoma? A dose-response meta-analysis of prospective cohort studies. Oncotarget. 2015;6: 13347-13358.
42. Kocher NJ, Rjepaj C, Robyak H, Lehman E, Raman JD. Hypertension is the primary component of metabolic syndrome associated with pathologic features of kidney cancer. World journal of urology. 2017;35: 67-72.
43. Liu X, Sun Q, Hou H, et al. The association between BMI and kidney cancer risk: An updated dose-response meta-analysis in accordance with PRISMA guideline. Medicine (Baltimore). 2018;97: e12860-e12860.
44. Linehan WM, Spellman PT, Ricketts CJ, et al. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. The New England journal of medicine. 2016;374: 135-145.

Supplementary Table 1.1 - Sex-adjusted Cox proportional hazard models with age as timescale for the association between Body Mass Index (BMI) and risk of clear-cell Renal Cell Carcinoma and papillary Renal Cell Carcinoma in the Netherlands Cohort Study on diet and cancer, 1986-2006

Characteristics	Subcohort	clear-cell Renal Cell Carcinoma			papillary Renal Cell Carcinoma		
		Cases	Sex-adjusted		Cases	Sex-adjusted	
	person-years	No.	HR	95% CI	No.	HR	95% CI
BMI at baseline (kg/m^2)							
<23	16723	55	1	Ref.	11	1	Ref.
23-<25	21298	90	1.18	(0.83-1.67)	18	1.04	(0.48-2.22)
25-<27	16372	81	1.36	(0.95-1.94)	10	0.70	(0.29-1.67)
≥27	15477	89	1.69	(1.19-2.40)	7	0.63	(0.24-1.65)
p trend			0.002			0.21	
Cont. (per kg/m^2)	69871	315	1.05	(1.01-1.09)	46	0.93	(0.84-1.03)
BMI at age 20 (kg/m^2) [a]							
<20.0	15384	61	1.42	(0.95-2.13)	6	0.82	(0.29-2.31)
20.0-<21.5	14576	45	1	Ref.	9	1	Ref.
21.5-<23	14558	67	1.49	(1.01-2.20)	10	1.11	(0.45-2.75)
≥23	15549	85	1.80	(1.24-2.62)	8	0.87	(0.34-2.25)
p trend			0.05			0.89	
Cont. (per kg/m^2)	60067	258	1.05	(1.00-1.10)	33	1.02	(0.90-1.14)
BMI change since age 20 (kg/m^2) [b]							
<1.5	14892	60	1	Ref.	8	1	Ref.
1.5-<3.5	16693	73	1.11	(0.77-1.61)	16	1.45	(0.57-3.69)
3.5-<5.5	13659	58	1.18	(0.79-1.75)	7	0.80	(0.27-2.37)
≥5.5	14823	67	1.41	(0.95-2.09)	2	0.22	(0.04-1.14)
p trend			0.08			0.02	
Cont. (per kg/m^2)	60067	258	1.04	(1.00-1.09)	33	0.89	(0.78-1.01)
Trouser size - Men							
<50	4745	26	1.14	(0.66-1.95)	8	1.92	(0.66-5.57)
50-<52	6736	32	1	Ref.	6	1	Ref.
52-<54	10137	63	1.29	(0.83-2.01)	12	1.25	(0.46-3.37)
54-<56	5553	41	1.51	(0.93-2.44)	8	1.54	(0.53-4.49)
≥56	2756	24	1.85	(1.06-3.23)	3	1.10	(0.27-4.43)
p trend			0.03			0.69	
Cont. (per size)	29928	186	1.06	(1.01-1.12)	37	1.00	(0.95-1.06)
Skirt size - women							
<42	7012	15	0.84	(0.44-1.63)	1	-	
42-<44	9372	24	1	Ref.	1	1	
44-<46	10272	31	1.18	(0.68-2.03)	3	-	
46-<48	6529	27	1.62	(0.93-2.85)	0	-	
≥48	3774	17	1.71	(0.90-3.24)	0	-	
p trend			0.01			-	
Cont. (per size)	36958	114	1.09	(1.03-1.15)	5	-	

[a] Analysis with BMI at age 20 have a restricted number of cases due to missing values
[b] Additionally adjusted for BMI at age 20

Supplementary Table 1.2 - Sex-adjusted Cox proportional hazard models with age as timescale for the association between cigarette smoking and risk of clear-cell Renal Cell Carcinoma and papillary Renal Cell Carcinoma in the Netherlands Cohort Study on diet and cancer, 1986-2006

Characteristics		clear-cell Renal Cell Carcinoma			papillary Renal Cell Carcinoma		
	Subcohort	Cases	Sex-adjusted		Cases	Sex-adjusted	
	person-years	No.	HR	95% CI	No.	HR	95% CI
Unadjusted for cigarette smoking frequency and duration							
Cigarette smoking status							
Never	27409	85	1	Ref.	10	1	Ref.
Former	24499	136	1.30	(0.95-1.77)	22	0.95	(0.43-2.12)
Current	17963	94	1.29	(0.93-1.78)	14	0.96	(0.41-2.12)
p trend			0.15			0.94	
Adjusted for cigarette smoking frequency and duration [a]							
Cigarette smoking status							
Never	27409	85	1	Ref.	10	1	Ref.
Former	24499	136	1.30	(0.94-1.80)	22	1.06	(0.45-2.46)
Current	17963	94	1.27	(0.91-1.78)	14	0.59	(0.22-1.61)
p trend			0.15			0.27	
Exclusively cigarette smoking							
Never	26090	73	1	Ref.	9	1	Ref.
Former	19102	93	1.36	(0.94-1.95)	10	0.87	(0.19-1.72)
Current	15348	82	1.50	(1.04-2.14)	13	0.41	(0.10-1.70)
p trend			0.02			0.22	
Solely in exclusive cigarette smokers							
Smoking frequency [b]							
>0-<20 cig/d	21689	102	1	Ref.	10	1	Ref.
≥20 cig/d	12761	73	1.06	(0.76-1.48)	13	1.50	(0.63-3.60)
Cont. (per 5 cig/d), former smoker	19102	93	1.05	(0.95-1.17)	10	1.08	(0.80-1.46)
Cont. (per 5 cig/d), current smoker	15348	82	1.03	(0.93-1.15)	13	1.11	(0.89-1.38)
Smoking duration [c]							
>0-<30 yrs	13701	65	1	Ref.	2	1	Ref.
≥30 yrs	20749	110	0.90	(0.62-1.32)	21	4.61	(1.10-19.27)
Cont. (per 5 yrs), former smoker	19102	93	1.02	(0.92-1.14)	10	1.67	(1.02-2.71)
Cont. (per 5 yrs), current smoker	15348	82	0.96	(0.85-1.08)	13	1.27	(0.71-2.28)
Number of years of smoking cessation [d]							
Current smoker	15348	82	1	Ref.	13	1	Ref.
>0-<15	10444	53	0.95	(0.67-1.34)	7	0.75	(0.30-1.87)
≥15	8657	40	0.91	(0.61-1.37)	3	0.44	(0.12-1.63)
Never smoker	26090	73	0.67	(0.48-0.94)	9	1.63	(0.53-5.01)
p trend, excl. never smokers			0.57			0.18	
Cont. (per 5 yrs), former smokers	19102	93	0.98	(0.87-1.11)	10	0.87	(0.58-1.31)

[a] Additionally adjusted for cigarette smoking duration (years, centered) and cigarette smoking frequency (cig/d, centered)

[b] Additionally adjusted for smoking duration (years, centered)

[c] Additionally adjusted for smoking frequency (cig/d, centered)

[d] Additionally adjusted for cigarette-years

Supplementary Table 1.3 - Sex-adjusted Cox proportional hazard models with age as timescale for the association between alcohol and risk of clear-cell Renal Cell Carcinoma and papillary Renal Cell Carcinoma in the Netherlands Cohort Study on diet and cancer, 1986-2006

Characteristics	Subcohort person-years	clear-cell Renal Cell Carcinoma			papillary Renal Cell Carcinoma		
		Cases	Sex-adjusted		Cases	Sex-adjusted	
		No.	HR	95% CI	No.	HR	95% CI
Alcohol consumption (g/d)							
Abstainer	16613	76	1	Ref.	7	1	Ref.
>0-<5	20731	79	0.79	(0.57-1.09)	13	1.37	(0.54-3.47)
≥5-<15	15589	64	0.73	(0.51-1.04)	14	1.37	(0.55-3.44)
≥15-<30	10810	63	0.95	(0.66-1.35)	8	0.91	(0.33-2.53)
≥30	6127	33	0.84	(0.54-1.30)	4	0.74	(0.21-2.61)
p trend			0.66			0.42	
Cont. (per 5 g/d)	69871	315	0.99	(0.96-1.03)	46	0.96	(0.87-1.05)
Alcohol consumption among stable abstainers/users (g/d) [a]							
Abstainer	12986	57	1	Ref.	7	1	Ref.
>0-<5	11867	46	0.84	(0.56-1.26)	6	0.82	(0.27-2.48)
≥5-<15	9435	42	0.88	(0.58-1.34)	6	0.82	(0.26-2.61)
≥15-<30	5832	38	1.20	(0.77-1.88)	3	0.56	(0.14-2.29)
≥30	3287	20	1.11	(0.64-1.93)	2	0.64	(0.12-3.51)
p trend			0.43			0.46	
Cont. (per 5 g/d)	43407	203	1.01	(0.97-1.06)	24	0.97	(0.84-1.13)

[a] Stable abstainers were defined as participants that reported no alcohol consumption 5 years before baseline. Stable users were defined as participants who reported that they drank equal amounts of beer or other alcoholic beverages 5 years before baseline.

Supplementary Table 1.4 - Sex-adjusted Cox proportional hazard models with age as timescale for the association between hypertension and risk of clear-cell Renal Cell Carcinoma and papillary Renal Cell Carcinoma in the Netherlands Cohort Study on diet and cancer, 1986-2006

Characteristics		clear-cell Renal Cell Carcinoma			papillary Renal Cell Carcinoma		
	Subcohort person-years	Cases	Sex-adjusted		Cases	Sex-adjusted	
		No.	HR	95% CI	No.	HR	95% CI
Self-reported hypertension diagnosis							
No	51868	216	1	Ref.	30	1	Ref.
Yes	18002	99	1.37	(1.07-1.76)	16	1.73	(0.94-3.19)
Use of antihypertensive medication							
No	56463	235	1	Ref.	35	1	Ref.
Yes	13408	80	1.51	(1.16-1.98)	11	1.43	(0.71-2.86)
Hypertension status & antihypertensive use							
No hyp or no med	59848	251	1	Ref.	35	1	Ref.
Hyp and med	10023	64	1.61	(1.20-2.15)	11	2.15	(1.09-4.24)

Abbreviations: hyp: Hypertension status, med: use of medication.

Supplementary Table 2.1 - Sex- and multivariable-adjusted Cox proportional hazard models with age as timescale for the association between Body Mass Index (BMI) and risk of Renal Cell Carcinoma (RCC) in the Netherlands Cohort Study on diet and cancer, 1986-2006

Characteristics	Renal Cell Carcinoma					
	Cases	Subcohort	Sex-adjusted		Multivariable-adjusted[a]	
	No.	person-years	HR	95% CI	HR	95% CI
BMI at baseline (kg/m^2)						
<23	98	16723	1	Ref.	1	Ref.
23-<25	151	21298	1.09	(0.83-1.43)	1.06	(0.80-1.39)
25-<27	120	16372	1.11	(0.83-1.47)	1.05	(0.78-1.40)
≥27	146	15477	1.56	(1.19-2.05)	1.46	(1.11-1.94)
p trend			0.002		0.009	
Cont. (per kg/m^2)	515	69871	1.05	(1.02-1.08)	1.04	(1.01-1.07)
BMI at age 20 (kg/m^2) [b]						
<20.0	104	15384	1.28	(0.94-1.74)	1.25	(0.92-1.71)
20.0-<21.5	86	14576	1	Ref.	1	Ref.
21.5-<23	109	14558	1.28	(0.95-1.73)	1.29	(0.95-1.75)
≥23	125	15549	1.39	(1.04-1.86)	1.35	(1.00-1.81)
p trend			0.28		0.29	
Cont. (per kg/m^2)	424	60067	1.03	(0.98-1.07)	1.02	(0.98-1.07)
BMI change since age 20 (kg/m^2) [c]						
<1.5	95	14892	1	Ref.	1	Ref.
1.5-<3.5	121	16693	1.13	(0.84-1.52)	1.10	(0.82-1.49)
3.5-<5.5	103	13659	1.27	(0.93-1.74)	1.23	(0.89-1.69)
≥5.5	105	14823	1.30	(0.94-1.80)	1.21	(0.86-1.69)
p trend			0.08		0.21	
Cont. (per kg/m^2)	424	60067	1.04	(1.01-1.08)	1.03	(0.99-1.07)
Trouser size - Men						
<50	44	4745	1.13	(0.74-1.74)	1.14	(0.74-1.75)
50-<52	55	6736	1	Ref.	1	Ref.
52-<54	108	10137	1.30	(0.91-1.84)	1.29	(0.91-1.83)
54-<56	69	5553	1.51	(1.03-2.21)	1.55	(1.05-2.28)
≥56	38	2756	1.67	(1.07-2.62)	1.64	(1.04-2.57)
p trend			0.02		0.02	
Cont. (per size)	314	29928	1.06	(1.02-1.10)	1.06	(1.01-1.10)
Corr. BMI/size			0.281			
Skirt size - women [d]						
<42	24	7012	0.83	(0.49-1.40)	0.83	(0.49-1.41)
42-<44	39	9372	1	Ref.	1	Ref.
44-<46	43	10272	1.00	(0.64-1.57)	1.00	(0.64-1.57)
46-<48	35	6529	1.30	(0.81-2.09)	1.22	(0.76-1.98)
≥48	33	3774	2.09	(1.28-3.41)	1.96	(1.18-3.27)
p trend			0.001		0.005	
Cont. (per size)	174	36958	1.10	(1.04-1.16)	1.09	(1.03-1.15)
Corr. BMI/size			0.771			

[a] Additionally adjusted for: smoking status (never/former/current), smoking duration (continuous, centered), smoking frequency (continuous, centered), pipe and/or cigar smoking (never/former/current), hypertension status (no/yes), alcohol consumption (g/d, continuous), diabetes (no/yes), models included a time-varying covariate for smoking frequency because of a potential violation of the proportional hazards assumption

[b] Analysis with BMI at age 20 have a restricted number of cases due to missing values

[c] Additionally adjusted for BMI at age 20

[d] Not adjusted for pipe and/or cigar smoking due to unstable estimates for these confounding factors

Supplementary Table 2.2 - Sex- and multivariable-adjusted Cox proportional hazard models with age as timescale for the association between cigarette smoking and risk of Renal Cell Carcinoma (RCC) in the Netherlands Cohort Study on diet and cancer, 1986-2006

Characteristics	Renal Cell carcinoma					
	Cases	Subcohort	Sex-adjusted		Multivariable-adjusted[a]	
	No.	person-years	HR	95% CI	HR	95% CI
Unadjusted for cigarette smoking frequency and duration						
Cigarette smoking status						
Never	136	27409	1	Ref.	1	Ref.
Former	219	24499	1.26	(0.98-1.62)	1.21	(0.93-1.57)
Current	160	17963	1.35	(1.04-1.74)	1.47	(1.13-1.91)
p trend			0.03		0.004	
Adjusted for cigarette smoking frequency and duration [b]						
Cigarette smoking status						
Never	136	27409	1	Ref.	1	Ref.
Former	219	24499	1.26	(0.97-1.64)	1.21	(0.91-1.6)
Current	160	17963	1.35	(1.03-1.77)	1.48	(1.12-1.94)
p trend			0.02		0.006	
Exclusively cigarette smoking						
Never	123	26090	1	Ref.	1	Ref.
Former	150	19102	1.19	(0.88-1.61)	1.21	(0.89-1.65)
Current	141	15348	1.46	(1.10-1.95)	1.54	(1.15-2.06)
p trend			0.01		0.004	
Solely in exclusive cigarette smokers						
Smoking frequency [c]						
>0-<20 cig/d	163	21689	1	Ref.	1	Ref.
≥20 cig/d	128	12761	1.16	(0.89-1.51)	1.12	(0.86-1.46)
Cont. (per 5 cig/d), former smoker	150	19102	1.07	(0.99-1.17)	1.06	(0.97-1.16)
Cont. (per 5 cig/d), current smoker	141	15348	1.07	(0.98-1.17)	1.07	(0.98-1.17)
Smoking duration [d]						
>0-<30 yrs	102	13701	1	Ref.	1	Ref.
≥30 yrs	189	20749	0.92	(0.68-1.25)	0.93	(0.68-1.26)
Cont. (per 5 yrs), former smoker	150	19102	1.02	(0.94-1.12)	1.03	(0.94-1.12)
Cont. (per 5 yrs), current smoker	141	15348	0.96	(0.86-1.07)	0.96	(0.86-1.07)
Number of years of smoking cessation [e]						
Current smoker	141	15348	1	Ref.	1	Ref.
>0-<15	90	10444	0.92	(0.71-1.2)	0.88	(0.67-1.14)
≥15	60	8657	0.77	(0.56-1.07)	0.74	(0.54-1.03)
Never smoker	123	26090	0.69	(0.53-0.90)	0.66	(0.50-0.86)
p trend, excl. never smokers			0.11		0.08	
Cont. (per 5 yrs), former smokers	150	19102	0.98	(0.89-1.08)	0.97	(0.88-1.07)

[a] Additionally adjusted for: pipe and/or cigar smoking (never/former/current), Body Mass Index (continuous), hypertension status (no/yes), alcohol consumption (g/d, continuous), diabetes (no/yes), models included a time-varying covariate for smoking frequency because of a potential violation of the proportional hazards assumption

[b] Additionally adjusted for cigarette smoking duration (years, centered) and cigarette smoking frequency (cig/d, centered)

[c] Additionally adjusted for smoking duration (years, centered)

[d] Additionally adjusted for smoking frequency (cig/d, centered)

[e] Additionally adjusted for cigarette-years

Supplementary Table 2.3 - Sex- and multivariable-adjusted Cox proportional hazard models with age as timescale for the association between alcohol and risk of Renal Cell Carcinoma (RCC) in the Netherlands Cohort Study on diet and cancer, 1986-2006

Characteristics	Renal Cell Carcinoma					
	Cases	Subcohort	Sex-adjusted		Multivariable-adjusted[a]	
	No.	person-years	HR	95% CI	HR	95% CI
Alcohol consumption (g/d)						
Abstainer	118	16613	1	Ref.	1	Ref.
>0-<5	135	20731	0.87	(0.67-1.13)	0.87	(0.66-1.13)
≥5-<15	104	15589	0.75	(0.56-0.99)	0.74	(0.56-0.99)
≥15-<30	101	10810	0.95	(0.71-1.28)	0.88	(0.65-1.19)
≥30	57	6127	0.89	(0.63-1.28)	0.82	(0.56-1.18)
p trend			0.63		0.28	
Cont. (per 5 g/d)	515	69871	1.00	(0.97-1.03)	0.99	(0.95-1.02)
Alcohol consumption among stable abstainers/users (g/d) [b]						
Abstainer	91	12986	1	Ref.	1	Ref.
>0-<5	73	11867	0.83	(0.60-1.15)	0.84	(0.60-1.16)
≥5-<15	62	9435	0.78	(0.55-1.10)	0.78	(0.54-1.11)
≥15-<30	60	5832	1.14	(0.79-1.65)	1.06	(0.71-1.56)
≥30	31	3287	1.02	(0.64-1.61)	0.92	(0.57-1.5)
p trend			0.63		0.95	
Cont. (per 5 g/d)	317	43407	1.01	(0.97-1.05)	0.99	(0.95-1.04)

[a] Additionally adjusted for: smoking status (never/former/current), smoking duration (continuous, centered), smoking frequency (continuous, centered), pipe and/or cigar smoking (never/former/current), hypertension status (no/yes), body mass index (kg/m2, continuous), diabetes (no/yes), models included a time-varying covariate for smoking frequency because of a potential violation of the proportional hazards assumption

[b] Stable abstainers were defined as participants that reported no alcohol consumption 5 years before baseline. Stable users were defined as participants who reported that they drank equal amounts of beer or other alcoholic beverages 5 years before baseline.

Supplementary Table 2.4: Sex- and multivariable-adjusted Cox proportional hazard models with age as timescale for the association between hypertension and risk of Renal Cell Carcinoma (RCC) in the Netherlands Cohort Study on diet and cancer, 1986-2006

Characteristics	Renal Cell Carcinoma					
	Cases	Subcohort	Sex-adjusted		Multivariable-adjusted[a]	
	No.	person-years	HR	95% CI	HR	95% CI
Self-reported hypertension diagnosis						
No	354	51868	1	Ref.	1	Ref.
Yes	161	18002	1.38	(1.13-1.69)	1.32	(1.08-1.62)
Use of antihypertensive medication						
No	391	56463	1	Ref.	1	Ref.
Yes	124	13408	1.40	(1.12-1.75)	1.20	(0.93-1.54)
Hypertension status & antihypertensive use						
No hyp or no med	410	59848	1	Ref.	1	Ref.
Hyp and med	105	10023	1.63	(1.29-2.07)	1.60	(1.12-2.28)

[a] Additionally adjusted for: smoking status (never/former/current), smoking duration (continuous, centered), smoking frequency (continuous, centered), pipe and/or cigar smoking (never/former/current), alcohol consumption (g/d, continuous), body mass index (kg/m2, continuous), diabetes (no/yes), models included a time-varying covariate for smoking frequency because of a potential violation of the proportional hazards assumption

CHAPTER 4

Kidney stones and the risk of renal cell carcinoma and upper tract urothelial carcinoma: the Netherlands Cohort Study

Jeroen AA van de Pol, Piet A van den Brandt and Leo J Schouten

Br. J. Cancer. 2019; 120: 368-374

Abstract

Background
We examined the association between kidney stones and renal cell carcinoma (RCC) and upper tract urothelial carcinoma (UTUC) risk in the Netherlands Cohort Study on diet and cancer.

Methods
In total, 120,852 participants aged 55-69 completed a self-administered questionnaire on diet, medical conditions and other risk factors for cancer at baseline (1986). After 20.3 years of cancer follow-up 4352 subcohort members, 544 RCC cases and 140 UTUC cases were eligible for case-cohort analysis. Hazard ratios (HR) and 95% confidence intervals (CIs) were estimated by multivariable-adjusted proportional hazards models.

Results
Kidney stones were associated with an increased RCC risk (HR: 1.39, 95%CI 1.06-1.85), vs. no kidney stones. Kidney stones were associated with an increased risk of papillary RCC (HR: 3.08, 95%CI 1.55-6.11), but not clear-cell RCC (HR: 1.14, 95%CI 0.79-1.65). UTUC risk was increased for participants with kidney stones (HR: 1.66, 95%CI 1.03-2.68). No heterogeneity of associations was found for UTUC in the ureter and renal pelvis. An early kidney stone diagnosis (≤40 years) was associated with an increased RCC and UTUC risk, compared to later diagnosis.

Conclusion
Kidney stones were associated with increased papillary RCC risk, but not clear-cell RCC risk. No heterogeneity was found for UTUC subtypes.

Introduction

Kidney stones, a common urological condition, affect five to ten percent of the population in Europe and North America [1]. Globally, the incidence and prevalence of kidney stones have increased over the years and are expected to increase further through the increasing prevalence of related medical conditions, such as obesity and diabetes mellitus[1, 2]. In general, kidney stone occurrences increase with age and are more common in men than in women[1, 3]. The likelihood of kidney stones decreases with an increased intake of fluids, fruits and vegetables. Sodium restriction also reduces the probability for kidney stones[3].

Several studies have assessed the relationship between kidney stones and renal cell carcinoma (RCC) and upper tract urothelial carcinoma (UTUC)[4-6]. Recently, a meta-analysis, based on eight case-control studies and one retrospective cohort study, found an increased risk of RCC and both ureter and renal pelvis cancer in individuals with kidney stones[7]. Furthermore, kidney stones were associated with an increased risk of RCC in males, but not in females. Three retrospective cohort studies not included in the aforementioned meta-analysis also found an increased risk of renal, ureter, or renal pelvis cancer in patients with urinary tract stones[8-10].

Increased cancer risks associated with kidney stones are commonly attributed to chronic inflammation and infections, which may lead to an altered proliferation in urothelial cells[5, 11, 12]. In turn, this process may lead to the development of a tumour. However, this association may also be explained by shared risk factors between kidney stones and RCC and UTUC[5, 11]. For example, obesity, diabetes mellitus, and several dietary factors are also associated with RCC risk[13, 14].

At present, solely case-control and retrospective cohort studies have assessed the relationship between kidney stones and RCC and UTUC risk. These study designs tend to be prone to information and selection bias, which may affect found associations. In addition, most of these studies were limited in their adjustment for confounding factors[7]. As a result, there is uncertainty whether kidney stones or a lifestyle related to kidney stone formation are associated with an increased RCC and UTUC risk. In this study, we investigate the relationship between self-reported history of kidney stones and the risk of RCC and UTUC in the Netherlands Cohort Study (NLCS) on Diet and Cancer. In the NLCS, detailed information on risk factors associated with kidney stones, RCC and UTUC has been collected prior to cancer development enabling this study to adjust for multiple confounders.

Methods

Study population:

The NLCS is a nation-wide prospective cohort study initiated in September 1986. It included 58 279 men and 62 573 women aged 55-69 years at baseline. The study design has been described in detail elsewhere[15]. In short, the study is a prospective cohort study initiated to investigate the association between diet and the development of cancer. For efficiency in data processing and analysis, a case-cohort design was used. A subcohort of 5,000 participants, of which 2,411 men and 2,589 women, was randomly sampled from the full cohort at baseline to estimate person-time at risk for the entire cohort[15].

At baseline, all participants completed a mailed, self-administered questionnaire on dietary habits and other risk factors for cancer. By filling in and returning the baseline questionnaire participants agreed to participate in the NLCS. Follow-up for cancer occurrence for all participants was conducted by computerised record linkage with the Netherlands Cancer Registry (NCR), the Netherlands Pathology Registry (PALGA), and causes of death registry maintained by Statistics Netherlands (CBS). In addition, subcohort members were followed up biannually for migration and vital status by contacting the participants and the municipalities. The completeness of cancer follow-up through record linkage is estimated to be at least 96%[16]. The institutional review boards of the Netherlands Organization for Applied Scientific Research TNO (Zeist) and Maastricht University (Maastricht) approved the NLCS. The NLCS was conducted in accordance with the Declaration of Helsinki.

In total, 608 RCC cases were identified in the NLCS between 1986 and 2006 (20.3 years). Histologically confirmed epithelial RCC cases were eligible for the collection of formalin-fixed paraffin-embedded (FFPE) tumour tissues. Overall, FFPE tumour tissues were collected for 454 (79.8%) of the eligible cases. Tumour histology was revised by two experienced pathologists according to the WHO-classification of RCC tumours[17]. Of the 454 RCC cases with available tumour tissue, 366 (80.6%) were clear-cell (cc)RCC cases, 60 (13.2%) papillary (p)RCC cases, 15 (3.3%) chromophobe RCC cases, and 13 (2.9%) other or undefined RCC cases. Further classification of pRCC cases resulted in 35 (7.7%) type 1 pRCC, 24 (5.3%) type 2 pRCC, and 1 (0.2%) undefined pRCC.

Cohort members with prevalent cancer at baseline, except skin cancer, and incomplete or inconsistent information on a priori selected confounders were excluded from analyses. Figure 1 shows the selection and exclusion of participants. In total, 544 RCC cases (International Classification of Diseases for Oncology 3 (ICD-O-3) C64.9), 140 UTUC cases, and 4352 subcohort members were included in this study. Of eligible RCC cases with confirmed tumour histology 332 were ccRCC cases and 48 were pRCC cases. Of UTUC cases 86 were renal pelvis cancer cases (ICD-O-3 C65.9) and 54 ureter cancer cases (ICD-O-3 C66.9).

Questionnaire data

All participants completed a mailed, self-administered questionnaire at baseline on dietary habits and other risk factors for cancer. The exposure to kidney stones was obtained from the question “Has a physician ever diagnosed kidney stones and what was your age at that time?”. Participants reported the age at first kidney stone diagnosis in 5-year increments starting from “younger than 30”. Information on dietary habits was obtained through a 150-item, semi-quantitative food frequency questionnaire (FFQ) focusing on habitual consumption of food and beverages during the year preceding baseline. Other risk factors included in the questionnaire considered for the association between RCC, UTUC and kidney stones were: age at baseline, anthropometry (height and weight), cigarette smoking (status, intensity, and duration), medical conditions (hypertension, diabetes mellitus and kidney stones including age at first diagnosis in 5-year increments), and the use of diuretic medication.

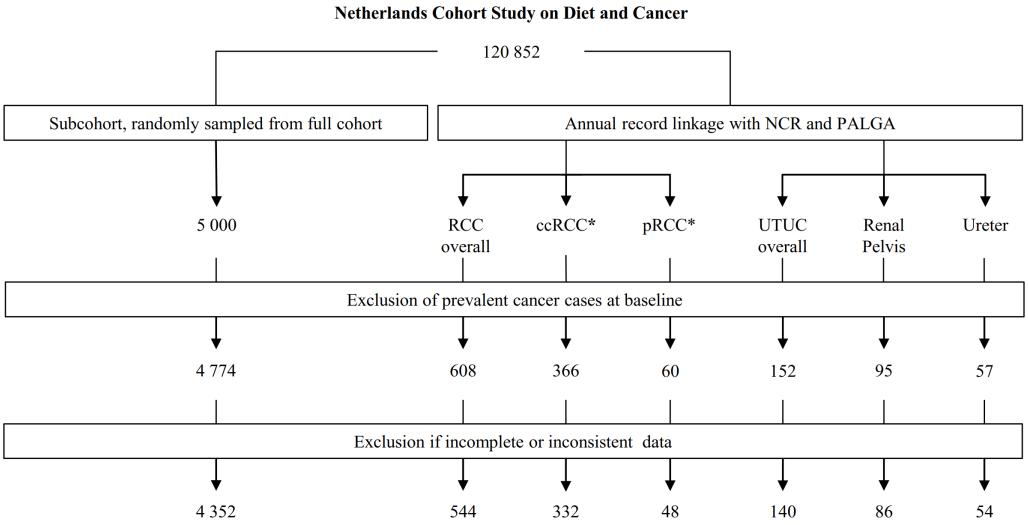

Figure 1 - Flow diagram of subcohort members and case subjects on whom the analyses were based. Abbreviations; RCC = renal cell carcinoma; ccRCC = clear cell renal cell carcinoma; pRCC - Papillary renal cell carcinoma; UTUC = upper tract urothelial carcinoma; NCR = Netherlands Cancer Registry; PALGA = the Netherlands Pathology Registry. * Histologically revised subtypes of RCC. Other subtypes were not assessed because of an insufficient number of cases.

Statistical analyses

Cox proportional hazards models were used to estimate age- and sex-adjusted and multivariable-adjusted hazard ratios (HR) and 95% confidence intervals (CIs). A priori confounders in the multivariable-adjusted model were body mass index (BMI, kg/m^2; continuous), hypertension (yes/no), smoking status (never, former and current), smoking intensity (cig/d, centered; continuous), and smoking duration (years, centered; continuous). Potential confounders were added to the multivariable-adjusted model if they affected the HR of kidney stones on RCC and UTUC risk for more than 10%. Potential confounders considered were diuretic medication use (yes/no), alcohol intake (g/d; continuous), fluid intake (liters; continuous), sodium intake (g/d, adjusted for total energy intake by residuals; continuous), total energy intake (kcal/d; continuous), fruit intake (g/d; continuous), vegetable intake (g/d; continuous) and family history of renal carcinoma (yes/no). None of these potential confounders affected the HR of kidney stones on RCC and UTUC risk by more than 10% and, therefore, none were included in the final multivariable-adjusted models.

Additionally, the relationship between kidney stones and the risk of histological RCC subtypes (ccRCC and pRCC), and UTUC subtypes based on location (renal pelvis cancer and ureter cancer) was analysed. Moreover, the association for age of first reported kidney stone diagnosis was assessed as the main exposure for RCC and UTUC risk. An age of 40 years was determined as cut-off point based on the median age at first kidney stone occurrence.

HRs and 95% CIs were obtained through Cox proportional hazards regression models using Stata statistical software: release 14 (StataCorp., 2015, College Station, TX). Person-years

at risk were calculated from baseline until registration of RCC or UTUC (depending on the cancer of interest) or until date of censoring by death, emigration, loss to follow-up or end of follow-up, whichever occurred first. During analyses for RCC or UTUC solely the cancer of interest was considered as an event and the development of other cancer types did not lead to censoring of observations. During analyses with histological RCC subtypes or UTUC subtypes based on location as the outcome, the first occurring subtype was considered as the time of censoring. The proportional hazards assumption was tested with Scaled Schoenfeld residuals and log-log curves[18]. Age at baseline was included as a time-varying covariate due to a violation of this assumption. Additional variance introduced due to the case-cohort approach by sampling a subcohort from the cohort was accounted for by use of the Huber-White sandwich estimator for standard errors, similar to the variance-covariance estimator by Barlow[19].
Tests for heterogeneity were performed to evaluate differences between histological subtypes and different localizations of tumours using the competing risks procedure in Stata. *P*-values were calculated based on a bootstrapping method that was developed for the case-cohort design. This procedure has been described in detail elsewhere[20, 21].
All tests were performed two-sided and *P* values <0.05 were considered statistically significant.

Results
Subcohort baseline characteristics categorised by history of kidney stones are presented in Table 1. In total, 8.4% of the subcohort had a history of kidney stones. In general, baseline characteristics for members with a history of kidney stones were similar to members without a history of kidney stones. Exceptions included a decreased proportion of current smokers in males with kidney stones and an increased fruit and vegetable intake in both men and women with kidney stones, when compared to participants without kidney stones in those categories.

Results from age- and sex-adjusted analyses did not differ substantially from the multivariable-adjusted results. Tables 2 and 3 present the results of the multivariable-adjusted analyses on history of kidney stones and age at first diagnosis of kidney stones for RCC and UTUC, respectively. In multivariable-adjusted analyses, participants with a history of kidney stones had a statistically significantly increased overall RCC risk (Table 2, HR: 1.39, 95% CI 1.06-1.85) compared to participants without a history of kidney stones. Kidney stones were significantly associated to pRCC (HR: 3.08, 95% CI 1.55-6.11), while no association was found for ccRCC (HR: 1.14, 95% CI 0.79-1.65). Tests for heterogeneity of associations for kidney stones indicated significant differences between ccRCC and pRCC for all participants (*P*=0.001). Type 1 pRCC risk and type 2 pRCC risk did not differ substantially from overall pRCC estimates (data not shown). In males, similar findings were obtained for RCC overall (HR: 1.42, 95% CI 1.04-1.93), pRCC (HR: 2.37, 95% CI 1.16-4.85) and ccRCC (HR: 1.20, 95% CI 0.80-1.81). In females, no statistically significant association was found for RCC overall (HR: 1.30, 95% CI 0.68-2.50) and ccRCC (HR: 0.88, 95% CI 0.35-2.22). An increased risk of pRCC (HR: 16.37, 95% CI 3.53-75.89) was found for females with a history of kidney stones compared to females without kidney stones. However, in females, the number of exposed cases was limited.

Table 1 - Characteristics of subcohort members categorized by history of kidney stones; Netherlands Cohort Study on Diet and Cancer, 1986-2006

Baseline characteristics[a]	Men		Women		Total	
	History of kidney stones		History of kidney stones		History of kidney stones	
	Yes	No	Yes	No	Yes	No
Number of persons, n %	267 (12.9)	1811 (87.2)	98 (4.3)	2176 (95.7)	365 (8.4)	3987 (91.6)
Exposures						
Age at first kidney stone, y	42.2 (11.3)	-	43.0 (13.5)	-	43.0 (10.9)	-
Time since first occurrence of kidney stones, y[b]	19.2 (11.5)	-	19.8 (12.9)	-	18.8 (10.9)	-
Covariates						
Age at baseline, y	60.9 (4.0)	61.3 (4.2)	62.3 (4.3)	61.5 (4.3)	61.3 (4.1)	61.4 (4.2)
BMI, kg/m^2	25.2 (2.8)	25.0 (2.6)	24.9 (3.3)	25.1 (3.6)	25.1 (2.9)	25.1 (3.1)
Cigarette smoking status, n %						
Never	39 (14.6)	250 (13.8)	58 (59.2)	1312 (60.3)	97 (26.6)	1562 (39.2)
Former	150 (56.2)	914 (50.5)	17 (17.4)	428 (19.7)	167 (45.8)	1342 (33.7)
Current	78 (29.2)	647 (35.7)	23 (23.5)	436 (20.0)	101 (27.7)	1083 (27.2)
Ever cigarette smokers only:						
Smoking duration, y	32.8 (11.6)	33.8 (11.7)	27.2 (14.0)	27.9 (12.3)	32.0 (12.1)	31.7 (12.2)
Smoking intensity, cig/d	17.4 (11.1)	17.1 (10.5)	13.0 (8.9)	11.7 (8.3)	16.7 (10.9)	15.2 (10.1)
Family history of renal cancer, n %	2 (0.8)	12 (0.7)	1 (1.0)	31 (1.4)	3 (0.8)	43 (1.1)
Hypertension, n %	67 (25.1)	425 (23.5)	31 (31.6)	641 (29.5)	98 (26.9)	1066 (26.7)
Diabetes Mellitus, n %	7 (2.6)	64 (3.5)	6 (6.1)	86 (4.0)	13 (3.6)	150 (3.8)
Use of diuretic medication, n %	29 (10.9)	140 (7.7)	15 (15.3)	309 (14.2)	44 (12.1)	449 (11.3)
Fluid intake, L/d	1.5 (0.5)	1.5 (0.5)	1.3 (0.4)	1.3 (0.4)	1.4 (0.5)	1.4 (0.4)
Alcohol intake, g/d[c]	17.1 (16.1)	17.6 (17.2)	8.1 (10.3)	8.7 (10.4)	15.0 (15.4)	13.4 (15.1)
Sodium intake, g/d	2.6 (0.9)	2.6 (0.9)	2.1 (0.7)	2.1 (0.7)	2.5 (0.9)	2.3 (0.8)
Fruit intake, g/d	157.3 (118.6)	153.7 (115.4)	196.5 (121.3)	194.6 (123.2)	167.8 (120.4)	176.1 (121.4)
Vegetable intake, g/d	194.1 (78.5)	190.9 (87)	202.0 (94.7)	192.8 (82.9)	196.2 (83.1)	191.9 (84.8)

[a] Data represent mean values (SD); for categorical variables N (%) is presented. Numbers may not add up to 100% as a result of missing values. Solely participants with complete information on the main exposures and the a priori selected confounders are presented in this table.
[b] Time between reported age of first kidney stone occurrence and age at baseline.
[c] In consumers only

A history of kidney stones was statistically significantly associated with an increased risk of UTUC overall in multivariable-adjusted models (Table 3, HR: 1.66, 95% CI 1.03-2.68). Similar estimates were found for specific UTUC localizations, namely the renal pelvis (HR: 1.76, 95% CI 0.96-3.23) and the ureter (HR: 1.50, 95% CI 0.71-3.18). Tests for heterogeneity of associations did not indicate significant differences between cancer of the renal pelvis and the ureter (P=0.841). Associations were similar for sex-specific estimates for UTUC overall and UTUC localizations. Due to the absence of exposed cases, no analyses on ureter cancer were performed in females.

An early diagnosis of kidney stones (<40 years), compared to a later kidney stone diagnosis (≥40 years), was statistically significantly associated with an increased overall RCC risk (HR: 2.10, 95% CI 1.21-3.65). For histological RCC subtypes a non-statistically significantly increased risk was found for both pRCC (HR: 3.52, 95% CI 0.95-13.01) and ccRCC (HR: 1.43, 95% CI 0.70-2.93). In UTUC analyses, an association was found for age (<40 years) at first diagnosis of kidney stones and overall UTUC risk (HR: 1.76, 95% CI 0.69-4.52). While no association was found for UTUC localized in the renal pelvis (HR: 1.13, 95% CI 0.30-4.22), an increased risk of ureter cancer was found (HR: 4.79, 95% CI 0.83-27.71). However, the number of participants was limited for this analysis, which reduced the power to find statistically significant results.

Discussion

In this study, an increased RCC and UTUC risk was found for participants with a history of kidney stones. Moreover, an increased pRCC risk, but not ccRCC risk, was observed in relation to a history of kidney stones. Furthermore, an increased RCC and UTUC risk was found for participants with a kidney stone diagnosis before their fortieth birthday. To our knowledge, this is the first prospective study to examine the relationship between kidney stones and RCC and UTUC risk and the first study to show heterogeneity of associations between pRCC and ccRCC.

The present study concurs with previously published studies on the relationship between kidney stones and RCC and UTUC risk. In a meta-analysis by Cheungpasitporn *et al*. an overall risk ratio of 1.76 (95% CI 1.24-2.49) was found comparing RCC risk for patients with kidney stones to those without kidney stones[7]. In the same meta-analysis a pooled risk ratio of 2.14 (95% CI 1.35-3.40) was found for transitional cell carcinoma, involving the ureter and renal pelvis[7]. Although HR estimates are lower in our study, we found similar associations for both RCC and UTUC risk. In addition, Cheungpasitporn *et al*. found an increased RCC risk associated with kidney stones in males, but not females[7]. In contrast, we found no difference between males and females.

To our knowledge, this is the first study to find an increased pRCC risk in participants with kidney stones. Nearly half of all pRCC cases in our study could be attributed to kidney stones based on the population attributable fraction using a multivariable-adjusted HR of 3.08 [22]. In general, pRCC is a heterogeneous RCC subtype consisting of two distinct subtypes characterised by genetic variations in the *MET-gene* for type 1 pRCC and in fumarate hydratase for type 2 pRCC[17, 23]. Previous studies on kidney stones have often been unable to assess the relationship with pRCC, either because pRCC has only been classified

Table 2 - Associations between kidney stones and risk of renal cell carcinoma subtypes; Netherlands Cohort Study on Diet and Cancer, 1986-2006

	Subcohort Person-time at risk (y)	RCC Overall[a]		ccRCC[a]		pRCC[a]		
		No. cases	HR (95% CI)	No. cases	HR (95% CI)	No. cases	HR (95% CI)	P. heterogeneity[f]
Unadjusted analyses:								
History of kidney stones:								
All participants:[b]								
No	67294	473	1 (reference)	296	1 (reference)	34	1 (reference)	
Yes	5982	71	1.40 (1.06-1.85)	36	1.15 (0.80-1.66)	14	3.20 (1.64-6.25)	0.001
Multivariable-adjusted analyses:								
History of kidney stones:								
All participants:[c]								
No	67294	473	1 (reference)	296	1 (reference)	34	1 (reference)	
Yes	5982	71	1.39 (1.05-1.84)	36	1.14 (0.79-1.65)	14	3.08 (1.55-6.11)	0.001
P for interaction[d]			0.815		0.593		0.022	
Males only:[e]								
No	28706	285	1 (reference)	172	1 (reference)	30	1 (reference)	
Yes	4285	60	1.42 (1.04-1.93)	31	1.20 (0.80-1.81)	11	2.37 (1.16-4.85)	0.084
Females only:[e]								
No	38588	188	1 (reference)	124	1 (reference)	4	1 (reference)	
Yes	1697	11	1.30 (0.68-2.50)	5	0.88 (0.35-2.22)	3	16.37 (3.53-75.89)	0.756
Age at first occurrence of kidney stones (participants with kidney stones only):								
All participants:[c]								
<40	2354	40	2.10 (1.21-3.65)	17	1.43 (0.70-2.93)	9	3.52 (0.95-13.01)	0.854
≥40	3553	31	1 (reference)	19	1 (reference)	5	1 (reference)	

[a] RCC: Renal Cell Carcinoma; Overall: no selection on morphological subtypes; ccRCC: clear cell Renal Cell Carcinoma; pRCC: papillary Renal Cell Carcinoma

[b] Model adjusted for age at baseline (y), sex (male/female), and age at baseline (y) as a time-varying covariate.

[c] Model adjusted for age at baseline (y), sex (male/female), body mass index (kg/m^2), smoking status (never, former, current), smoking intensity (cig/d, centered), smoking duration (y, centered), hypertension (yes, no), and age at baseline (y) as a time-varying covariate.

[d] Test on whether sex modifies the (multivariable-adjusted) relationship between the history of kidney stones and RCC risk, based on the Wald test for the interaction term.

[e] Model adjusted for age at baseline (y), body mass index (kg/m^2), smoking status (never, former, current), smoking intensity (cig/d, centered), smoking duration (y, centered), hypertension (yes, no), and age at baseline (y) as a time-varying covariate.

[f] Test for heterogeneity between histological RCC subtypes

Table 3 - Associations between kidney stones and risk of upper tract urothelial carcinoma localizations; Netherlands Cohort Study on Diet and Cancer, 1986-2006

	Subcohort person-time at risk (y)	UTUC Overall[a]		Renal Pelvis		Ureter		
		No. cases	HR (95% CI)	No. cases	HR (95% CI)	No. cases	HR (95% CI)	P. heterogeneity[f]
Unadjusted analyses:								
History of kidney stones:								
All participants:[b]								
No	67389	118	1 (reference)	72	1 (reference)	46	1 (reference)	
Yes	5981	22	1.64 (1.02-2.64)	14	1.74 (0.95-3.16)	8	1.49 (0.70-3.16)	0.841
Multivariable-adjusted analyses:								
History of kidney stones:								
All participants:[c]								
No	67389	118	1 (reference)	72	1 (reference)	46	1 (reference)	
Yes	5981	22	1.66 (1.03-2.68)	14	1.76 (0.96-3.23)	8	1.50 (0.71-3.18)	0.841
P for interaction[d]			0.923		0.395		-[g]	
Males only:[e]								
No	28774	81	1 (reference)	49	1 (reference)	32	1 (reference)	
Yes	4284	19	1.65 (0.98-2.79)	11	1.59 (0.81-3.14)	8	1.74 (0.80-3.80)	0.835
Females only:[e]								
No	38615	37	1 (reference)	23	1 (reference)	14	1 (reference)	
Yes	1697	3	1.70 (0.51-5.72)	3	2.77 (0.80-9.57)	0	-[g]	-
Age at first occurrence of kidney stones (participants with kidney stones only):								
All participants:[c]								
<40	2354	11	1.76 (0.69-4.52)	5	1.13 (0.30-4.22)	6	4.79 (0.83-27.71)	0.892
≥40	3552	11	1 (reference)	9	1 (reference)	2	1 (reference)	

[a] UTUC: Upper Tract Urothelial Carcinoma; Overall: no selection on subtypes by localization

[b] Model adjusted for age at baseline (y), sex (male/female), and age at baseline (y) as a time-varying covariate.

[c] Model adjusted for age at baseline (y), sex (male/female), body mass index (kg/m^2), smoking status (never, former, current), smoking intensity (cig/d, centered), smoking duration (y, centered), hypertension (yes, no), and age at baseline (y) as a time-varying covariate.

[d] Test on whether sex modifies the (multivariable-adjusted) relationship between the history of kidney stones and UTUC risk, based on the Wald test for the interaction term.

[e] Model adjusted for age at baseline (y), body mass index (kg/m^2), smoking status (never, former, current), smoking intensity (cig/d, centered), smoking duration (y, centered), hypertension (yes, no), and age at baseline (y) as a time-varying covariate.

[f] Test for heterogeneity between renal pelvis and ureter subtypes

[g] Insufficient number of exposed cases

as a distinct tumour type since 1996, or because they did not contain information on tumour histology[24]. Our study was able to assess tumour histology through centralised revision by two pathologists. Even though pRCC is a very heterogeneous subtype of RCC, we did not find differences between type 1 pRCC and type 2 pRCC.

There is uncertainty regarding the biological mechanism that may relate kidney stones to kidney cancer. Kidney stones are presumed to cause chronic irritation in the local environment of the kidney and ureter[5-7]. In general, chronic irritation and infection recruit inflammatory cells, which secrete cytokines and chemokines. In turn, free radical species from oxygen and nitrogen are produced, facilitating the onset of cancer through, among others, increased cell proliferation[12]. However, more insight on the role of kidney stones in this process is needed to elucidate the found associations.

Chow *et al*. found that most renal pelvis and ureter cancers occurred on the same side as kidney stone formation, which could indicate that kidney stones are exerting the effect in UTUC[5]. In animal studies, induced stone formation was correlated to the development of bladder cancer[25, 26]. By suppressing stone formation in rats, the effect of kidney stones could be attributed to the irritative stimulation by kidney stones, rather than to metabolites of the stone inducing factor[25]. However, both ccRCC and pRCC are thought to originate from the proximal convoluted renal tubule[23]. It is deemed unlikely that stones or stone-forming crystals deposit in the proximal convoluted renal tubule. Kidney stones tend to form in locations where there is a combination of supersaturation of the urine and where there is a change in the luminal diameter of the renal tubules, such as the loop of Henle, the distal tubules and in the collecting ducts[27]. Therefore, urinary solutes or a predisposing lifestyle, rather than actual stone formation in the kidney, might play a role in the development of these cancer subtypes. In contrast to pRCC, ccRCC risk was not associated with a history of kidney stones in this study. In general, genetic susceptibility and the interaction with environmental exposures are believed to influence RCC risk[28]. Hypothetically, tumour development could be related to the presence of stone-forming salts in the filtrate of the proximal tubules. The presence of these solutes may affect cell metabolism, which could potentially result in the development of distinct renal cancer subtypes.

UTUC risk was increased in participants with kidney stones, compared to participants without kidney stones, but no difference was found between the localization in the renal pelvis or the ureter. In contrast to the proximal tubule, stone formation is common in the renal pelvis and ureter, which enables kidney stones to cause chronic irritation and inflammation to urothelial cells. In turn, this may explain the increased UTUC risk in relation to kidney stones.
In this study, an age below 40 years at first kidney stone diagnosis was potentially associated with an increased RCC and UTUC risk. However, further research on this potential association is needed as the number of cases eligible for these analyses was limited. An earlier kidney stones diagnosis could provide a longer time period for kidney stones to induce chronic irritation to the local environment or for potentially harmful solutes in the urine to have a carcinogenic effect. Therefore, the found associations could indicate that the lifestyle of kidney stone formers may already play a role in the development of cancer during the early stages of life.

The strengths of the present study were the complete follow-up, the extensive information on potential confounders and the differentiation between histological subtypes based on the centralised revision by experienced pathologists.

However, this study was also subject to limitations. Information on kidney stones was retrieved from a self-administered questionnaire at baseline. Consequentially, kidney stone occurrences beyond the age at baseline may have been missed for participants with and without cancer. However, peak kidney stone incidence is expected at 40 to 49 years of age. Therefore, effects on our results are assumed to be limited. Furthermore, information obtained through self-reported questionnaires may contain inaccuracies regarding the diagnosis of kidney stones. However, the prevalence and incidence of kidney stones were as expected in the population and the RCC and UTUC risk was the greatest before 40 years of age[1]. Therefore, we think that our results are generalizable for the population. In this study, we did not have information on kidney stone composition, frequency and laterality. This information could provide additional insight on the mechanisms behind the found association in future studies. Residual confounding could have affected our results. However, as all models were extensively adjusted for confounders, we expect this effect to be limited. Lastly, a diagnosis of kidney stones may warrant additional surveillance, which could lead to an earlier detection of RCC. However, in our study, the average tumour size was larger in cases with a history of kidney stones, compared to cases without a history of kidney stones (72 mm vs. 65 mm, respectively), which makes bias due to earlier detection unlikely.

In light of the findings of this study, more research is needed to unravel the mechanisms behind the relation of kidney stones and RCC and UTUC. Firstly, future studies are required to ascertain the relationship between kidney stones and pRCC. Secondly, more studies are needed on kidney stone composition, stone laterality and exposure to stone-forming solutes to uncover the impact on cell metabolism and cancer development. Lastly, more studies are required to get a better insight on sex-specific differences in RCC and UTUC risk as a result of kidney stones.

References

1. Romero V, Akpinar H, Assimos DG. Kidney stones: a global picture of prevalence, incidence, and associated risk factors. Reviews in urology. 2010;12: e86-96.
2. Sfoungaristos S, Gofrit ON, Yutkin V, Pode D, Duvdevani M. Prevention of renal stone disease recurrence. A systematic review of contemporary pharmaceutical options. Expert opinion on pharmacotherapy. 2015;16: 1209-1218.
3. Moe OW. Kidney stones: pathophysiology and medical management. Lancet (London, England). 2006;367: 333-344.
4. McCredie M, Stewart JH. Risk factors for kidney cancer in New South Wales, Australia. II. Urologic disease, hypertension, obesity, and hormonal factors. Cancer causes & control : CCC. 1992;3: 323-331.
5. Chow WH, Lindblad P, Gridley G, et al. Risk of urinary tract cancers following kidney or ureter stones. Journal of the National Cancer Institute. 1997;89: 1453-1457.
6. Chung SD, Liu SP, Lin HC. A population-based study on the association between urinary calculi and kidney cancer. Canadian Urological Association journal = Journal de l'Association des urologues du Canada. 2013;7: E716-721.
7. Cheungpasitporn W, Thongprayoon C, O'Corragain OA, et al. The risk of kidney cancer in patients with kidney stones: a systematic review and meta-analysis. QJM : monthly journal of the Association of Physicians. 2015;108: 205-212.
8. Sun LM, Lin CL, Chang YJ, et al. Urinary tract stone raises subsequent risk for urinary tract cancer: a population-based cohort study. BJU international. 2013;112: 1150-1155.
9. Shih CJ, Chen YT, Ou SM, Yang WC, Chen TJ, Tarng DC. Urinary calculi and risk of cancer: a nationwide population-based study. Medicine. 2014;93: e342.
10. Lin CL, Huang WT, Fan WC, et al. Associations between interventions for urolithiasis and urinary tract cancer among patients in Taiwan: The effect of early intervention. Medicine. 2016;95: e5594.
11. Schlehofer B, Pommer W, Mellemgaard A, et al. International renal-cell-cancer study. VI. the role of medical and family history. International journal of cancer. 1996;66: 723-726.
12. Federico A, Morgillo F, Tuccillo C, Ciardiello F, Loguercio C. Chronic inflammation and oxidative stress in human carcinogenesis. International journal of cancer. 2007;121: 2381-2386.
13. Stewart JH, Buccianti G, Agodoa L, et al. Cancers of the kidney and urinary tract in patients on dialysis for end-stage renal disease: analysis of data from the United States, Europe, and Australia and New Zealand. Journal of the American Society of Nephrology : JASN. 2003;14: 197-207.
14. McLaughlin JK, Lipworth L, Tarone RE, Blot WJ. Renal Cancer. In: Schottenfeld D, Fraumeni JF, editors. Cancer Epidemiology and Prevention. New York: Oxford University Press, 2006:1087-1100.
15. van den Brandt PA, Goldbohm RA, van 't Veer P, Volovics A, Hermus RJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. Journal of clinical epidemiology. 1990;43: 285-295.
16. Goldbohm RA, van den Brandt PA, Dorant E. Estimation of the coverage of dutch minicipalities by cancer registries and PALGA based on hospital discharge data. Tijdschr Soc Gezondheidsz. 1994;72: 80-84.
17. Eble JN, Sauter G, Epstein JI, Sesterhenn IA. World Health Organization classification of Tumours. Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs. Lyon: IARC Press, 2004.
18. Schoenfeld D. Partial Residuals for The Proportional Hazards Regression Model. Biometrika.

1982;69: 239-241.
19. Barlow WE. Robust variance estimation for the case-cohort design. Biometrics. 1994;50: 1064-1072.
20. de Vogel S, Bongaerts BW, Wouters KA, et al. Associations of dietary methyl donor intake with MLH1 promoter hypermethylation and related molecular phenotypes in sporadic colorectal cancer. Carcinogenesis. 2008;29: 1765-1773.
21. Wacholder S, Gail MH, Pee D, Brookmeyer R. Alternative Variance and Efficiency Calculations for the Case-Cohort Design. Biometrika. 1989;76: 117-123.
22. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. American journal of public health. 1998;88: 15-19.
23. Shuch B, Amin A, Armstrong AJ, et al. Understanding pathologic variants of renal cell carcinoma: distilling therapeutic opportunities from biologic complexity. European urology. 2015;67: 85-97.
24. Delahunt B, Eble JN, McCredie MR, Bethwaite PB, Stewart JH, Bilous AM. Morphologic typing of papillary renal cell carcinoma: comparison of growth kinetics and patient survival in 66 cases. Human pathology. 2001;32: 590-595.
25. Ogasawara H, Imaida K, Ishiwata H, et al. Urinary bladder carcinogenesis induced by melamine in F344 male rats: correlation between carcinogenicity and urolith formation. Carcinogenesis. 1995;16: 2773-2777.
26. Okumura M, Hasegawa R, Shirai T, Ito M, Yamada S, Fukushima S. Relationship between calculus formation and carcinogenesis in the urinary bladder of rats administered the non-genotoxic agents thymine or melamine. Carcinogenesis. 1992;13: 1043-1045.
27. Khan SR. Renal tubular damage/dysfunction: key to the formation of kidney stones. Urological research. 2006;34: 86-91.
28. Chow WH, Dong LM, Devesa SS. Epidemiology and risk factors for kidney cancer. Nature reviews. Urology. 2010;7: 245-257.

CHAPTER 5

Germline polymorphisms in the Von Hippel-Lindau and Hypoxia-inducible factor 1-alpha genes, gene-environment and gene-gene interactions and renal cell cancer

Jeroen AA van de Pol, Piet A van den Brandt, Manon van Engeland, Roger WL Godschalk, Frederik-Jan van Schooten, Janneke GF Hogervorst and Leo J Schouten

Adapted from *Sci Rep.* 2020; 10:137

Abstract

We investigated the relationship between germline single nucleotide polymorphisms (SNPs) in *Von Hippel-Lindau (VHL)* and *Hypoxia-inducible factor 1-alpha (HIF1A),* and their gene-environment and gene-gene interactions, and clear-cell RCC (ccRCC) risk. Furthermore, we assessed the relationship between *VHL* SNPs and *VHL* promoter methylation. Three *VHL* polymorphisms and one *HIF1A* polymorphism were genotyped in the Netherlands Cohort Study. In 1986, 120,852 participants aged 55-69 completed a self-administered questionnaire on diet and lifestyle and toenail clippings were collected. Toenail DNA was genotyped using the Sequenom MassARRAY platform. After 20.3 years, 3004 subcohort members and 406 RCC cases, of which 263 ccRCC cases, were eligible for multivariate case-cohort analyses. *VHL*_rs779805 was associated with RCC (Hazard Ratio (HR) 1.53; 95% Confidence Interval (CI) 1.07-2.17) and ccRCC risk (HR 1.88; 95% CI 1.25-2.81). No associations were found for other SNPs. Potential gene-environment interactions were found between alcohol consumption and selected SNPs. However, none remained statistically significant after multiple comparison correction. No gene-gene interactions were observed between *VHL* and *HIF1A*. *VHL* promoter methylation was not associated with *VHL* SNPs. *VHL* SNPs may increase (cc)RCC susceptibility. No associations were found between gene-environment and gene-gene interactions and (cc)RCC risk and between *VHL* promoter methylation and *VHL* SNPs.

Introduction:
Genetic and epigenetic alterations in the *Von Hippel-Lindau* (VHL) gene are important drivers of carcinogenesis in clear-cell renal cell carcinoma (ccRCC)[1]. For sporadic ccRCC, biallelic inactivation of *VHL* because of rare, but highly penetrant, somatic mutations is relatively common[2, 3]. Previous studies have estimated that 50-82% of patients with sporadic ccRCC have a mutation in the *VHL* gene[4-8] The *VHL* gene encodes the VHL tumor suppressor protein (pVHL). Inactivation of pVHL leads to the unchecked accumulation of hypoxia-inducible factor 1 alpha (HIF1A), which facilitates oxygen delivery, adaptation to oxygen deprivation and angiogenesis[1, 9]. Therefore, genetic or epigenetic alterations in *VHL* and *HIF1A* may lead to enhanced cell survival and carcinogenesis.

In contrast to the rare, but highly penetrant, sequence alterations leading to functional *VHL* loss, some germline Single Nucleotide Polymorphisms (SNPs) are highly frequent, but have a low penetrance. In general, SNPs account for many different phenotypes as they may alter disease susceptibility by affecting the gene's function[10]. Genome-wide association studies (GWAS) have not found an association with *VHL* and *HIF1A* loci[11-18]. However, candidate gene studies have found conflicting evidence on the relationship between *VHL* SNPs and (cc) RCC risk, with some studies indicating a positive association[19, 20], while others indicate no association [21]. In previous studies, *HIF1A* SNPs have been associated with RCC prognosis, but not with (cc)RCC development[21, 22].

Previous studies have indicated the importance of assessing the interplay between genetic, epigenetic and environmental triggers when assessing ccRCC risk. Moore *et al.* found increased promoter hypermethylation in sporadic ccRCC when certain *VHL* polymorphisms were present[23]. In addition, multiple studies have indicated potential gene-environment interactions between germline SNPs and environmental factors in RCC[24-28]. To our knowledge, the relationship between established environmental risk factors associated with RCC risk, namely smoking, hypertension, obesity and alcohol consumption[29], and *VHL* and *HIF1A* SNPs remains unstudied.

Therefore, we investigated the relationship between three selected germline *VHL* SNPs and one *HIF1A* SNP and (cc)RCC risk in the Netherlands Cohort Study on diet and cancer (NLCS). In addition, interactions between *VHL* and *HIF1A* SNPs and smoking, hypertension, body mass index (BMI) and alcohol consumption were studied. Lastly, we investigated the association between *VHL* promoter methylation and *VHL* SNPs.

Methods:
Study design
The NLCS is a nation-wide prospective cohort study initiated in September 1986 with the inclusion of 120,852 participants aged 55-69 years to study the relationship between diet and cancer. The study design has been described in detail elsewhere[30]. In short, a case-cohort design was used for efficiency in data processing and follow-up for vital status. Cases were derived from the entire cohort, whereas a subcohort of 5000 participants, consisting of 2411 men and 2589 women, was randomly sampled at baseline to estimate person years at risk for the entire cohort. The subcohort was followed up biennially for migration and vital status information by contacting participants and using computerized municipalities registries.

Using the subcohort, person-years at risk were calculated from baseline until registration of RCC, or until date of censoring by death, emigration, loss to follow-up or end of follow-up, whichever occurred first. Cancer follow-up for the full cohort was conducted by computerized record linkage with the Netherlands Cancer Registry (NCR), the Netherlands Pathology Registry (PALGA), and causes of death registry maintained by Statistics Netherlands (CBS)[31]. Follow-up for vital status of the subcohort was nearly 100% complete after 20.3 years. The completeness of cancer follow-up is estimated to be over 96%[32].

Individuals with prevalent cancer, excluding skin cancer, at baseline were excluded. After 20.3 years of follow-up, 608 RCC cases were identified (International Classification of Diseases for Oncology 3 (ICD-O-3):C64). Histologically confirmed epithelial RCC cases were eligible for the collection of formalin-fixed paraffin-embedded (FFPE) tumor tissue. Tumor blocks were collected for 454 out of 568 eligible cases (80%). Two experienced pathologists revised the tumor histology according to the WHO-classification of RCC tumors[33]. Based on this revision 366 (81%) of the cases with available tumor blocks were classified as ccRCC cases, 60 (13%) as papillary RCC cases, 15 (3.3%) chromophobe RCC cases, and 13 (2.9%) other or undefined RCC cases.

Ethics Statement

Individuals invited to participate in the NLCS received an invitation letter with details on the study and the use of their data. In addition, they received the baseline questionnaire, which included an envelope for returning toenail clippings. By completing and returning the baseline questionnaire, individuals consented to participate in the NLCS (response rate 35.5%). Individuals were informed about the possibility to end their participation at any time, at which point all their data would be removed. All methods were performed in accordance with the relevant guidelines and regulations that were applicable at that time (1986). The institutional review boards of Maastricht University (Maastricht) and the Netherlands Organization for Applied Scientific Research TNO (Zeist) approved the NLCS (February 2, 1985 and January 6, 1986, respectively). The institutional review board of Maastricht University (Maastricht) later re-evaluated the original approval of the study protocol and procedures (2010). Based on the re-evaluation the institutional review board amended the original approval to include the genotyping of SNPs (April 12, 2010). Participants did not provide written informed consent to the sharing of data.

Gene and SNP selection

Genes and SNPs related to RCC risk were selected through literature search. Priority was given to SNPs with a MAF $\geq$ 20% in Caucasians and primers had to be compatible with RAAS-pathway SNPs present on the multiplex assay[34]. Consequently, three *VHL* SNPs (rs779805, rs265318 and rs1642739) and one *HIF1A* SNP (rs2301111) were selected. All included *VHL* SNPs were selected based on their association with *VHL* promoter methylation in previous research[23]. The included *HIF1A* tag-SNP had the MAF of the *HIF1A* SNPs compatible with the assay.

Tissue collection and DNA isolation

Approximately 90,000 participants provided toenail clippings at baseline, which have been shown to be a valid source of DNA for the genotyping of germline genetic variants[35]. DNA

was isolated according to the DNA isolation protocol by Cline *et al* [36]. To increase the number of cases with available DNA, DNA was isolated from FFPE healthy tissue, as described by van Houwelingen *et al.*[37], for 67 RCC cases without toenail clippings. There were no substantial quality differences between DNA samples from toenail and FFPE healthy tissue [34]. In total, 3582 (75%) subcohort members and 502 (83%) RCC cases were genotyped.
SNP genotyping was performed on the Sequenom MassARRAY platform using the iPLEX assay (Sequenom Inc., Hamburg, Germany), as described previously[34]. This method provides suitable SNP call rates and reproducibility using toenail DNA[35].

DNA methylation of the CpG island of the *VHL* gene promoter region, of which methylation has been associated with inhibition of *VHL* gene expression[38], in RCC tumor blocks was determined by chemical modification of genomic DNA with sodium bisulfite and subsequent methylation-specific PCR analysis (MSP) as previously described elsewhere[39-41]. MSP primer design was based on the MBD-affinity massive parallel sequencing data. Detailed information on primer sequences and MSP conditions are available elsewhere[24].

Questionnaire information
All participants completed a mailed, self-administered, questionnaire on diet and other cancer risk factors for cancer at baseline (1986)[42]. Information on dietary habits was obtained through a 150-item, semi-quantitative food frequency questionnaire (FFQ) focusing on habitual consumption of food and beverages during the year preceding baseline.

Cigarette smoking status, frequency and duration were based on self-reported information. Participants reported hypertension as diagnosed by a physician, preceding baseline. Participants were asked to report the use of any drugs that they used longer than 6 months. From this information, the use of antihypertensive medication was extracted. BMI was calculated using self-reported height and weight from the baseline questionnaire. Questions on beer, red wine, white wine, sherry, fortified wines, liqueur, and liquor were used to assess the consumption of alcohol. Participants who consumed alcoholic beverages less than once a month were considered non-users. Standard glass sizes were defined as 200 ml for beer, 105 ml for wine, 80 ml for sherry, and 45 ml for both liqueur and liquor[43]. These values corresponded to 8, 10, 11, 7 and 13 grams of alcohol, respectively. Mean daily alcohol consumption was calculated by multiplying the consumption frequency and the standardized item unit.

Statistical analyses
Cox proportional hazards models were used to estimate age- and sex-adjusted and multivariable-adjusted hazard ratios (HR) and 95% confidence intervals (CIs). *A priori* selected covariables in the multivariable-adjusted model were BMI (kg/m^2, continuous), hypertension (yes,no), cigarette smoking status (never, former, current), intensity (cig/d, centered; continuous), duration (years, centered; continuous) and alcohol consumption (g/d, continuous).

The most common allele was used as the reference allele. Associations between genotypes and RCC and ccRCC risk were assessed using additive and dominant models. Results of SNPs with a MAF<0.25 were interpreted using a dominant model for power reasons. SNP

allele frequencies in the subcohort were tested against departure from the Hardy-Weinberg Equilibrium using the Pearson χ^2-test, as calculated with the Stata program 'hwsnp'[44]. Gene-environment interactions were tested with the Wald χ^2 test. Gene-environment analyses were adjusted for multiple comparisons with the adaptive Benjamini-Hochberg false discovery rate (FDR) procedure with a q-value threshold of 10%[45]. Sensitivity analyses were performed to explore the impact of using alternative categorizations for BMI (<20 kg/m2, 20-<25 kg/m2, 25-<30 kg/m2 and 30+ kg/m2), smoking status (never, ever), hypertension (no self-reported hypertension or no self-reported antihypertensive medication, hypertension with self-reported hypertensive medication) and alcohol consumption (0 g/d, 0.1-4 g/d, 5-14 g/d, 15-29 g/d and 30+ g/d) when assessing gene-environment interactions. Gene-gene interactions between *VHL* SNPs and the selected *HIF1A* SNP were tested using the Wald χ^2 test. In a case-only analysis, the association between *VHL* SNPs and VHL tumor promoter methylation status (methylated, unmethylated) was assessed using multiple logistic regression for both RCC and ccRCC.

All analyses were performed using Stata Statistical Software: Release 15 (StataCorp., 2017, College Station, TX). The proportional hazards assumption was tested using scaled Schoenfeld residuals[46]. A violation of the assumption was apparent for age. Therefore, all models were adjusted for age as a time-dependent covariable. With the exception of FDR-corrected analyses, a p-value <0.05 was considered statistically significant.

Results

After excluding participants with missing values for predefined confounders 3004 subcohort members and 406 RCC cases, of which 263 ccRCC cases, were included in the analyses. The proportion of men was higher in both RCC and ccRCC cases when compared to the subcohort (Table 1). In addition, cases were more often smokers and were more often diagnosed with hypertension when compared to the subcohort.

Genotype and allele frequencies for the four selected SNPs in subcohort members of the NLCS are presented in Supplementary Table 1. All selected SNPs adhered to the Hardy-Weinberg Equilibrium. Only *VHL*_rs779805 had a minor allele frequency (MAF) above 25% and is, therefore, assessed primarily using additive models.

Main SNP effects

In both age- and sex-adjusted analyses and multivariable-adjusted analyses, an association with (cc)RCC risk was observed for SNPs in *VHL*_rs779805, but not for SNPs in *VHL*_rs1642739, *VHL*_rs265318 and *HIF1A*_rs2301111 (Table 2). In multivariable-adjusted analyses individuals carrying the AG (vs. AA) genotype of *VHL*_rs779805 had a statistically significantly increased RCC risk (HR 1.32, 95%CI 1.06-1.66), and the GG (vs. AA) genotype was associated with a statistically significantly increased RCC risk (HR 1.53, 95%CI 1.07-2.17). In addition, a statistically significant per-allele *p* for trend was observed (p=0.004). In multivariable-analyses for ccRCC risk, the AG (vs. AA) genotype for *VHL*_rs779805 was associated with a statistically significantly increased ccRCC risk (HR 1.35, 95% CI 1.02-1.78), as was the GG (vs. AA) genotype of *VHL*_rs779805 (HR 1.88, 95%CI 1.25-2.81).

Table 1 - Baseline characteristics of the subcohort and renal cell carcinoma (RCC) and clear cell renal cell carcinoma (ccRCC) cases; Netherlands Cohort Study on diet and cancer, 1986-2006

Baseline characteristics (mean (SD))	Subcohort members	RCC	ccRCC
Total (n)	3004	406	263
Age (y)	61.3 (4.2)	60.7 (3.9)	60.6 (3.9)
Male sex (%)	49.6	65.8	64.3
Tumor stage (%)[a]			
Stage 1/2	-	49.8	51.0
Stage 3/4	-	38.2	39.5
Undefined	-	12.1	9.5
Cigarette smoking status (%)			
Never smoker	36.8	26.9	26.6
Former smoker	36.4	43.8	45.3
Current smoker	26.9	29.3	28.1
Smoking intensity (cig/d)[b]	15.2 (10.2)	17.0 (11.7)	16.2 (10.5)
Smoking duration (y)[b]	31.5 (12.2)	32.0 (11.6)	31.4 (11.3)
Hypertension (%)	26.4	33.5	33.5
BMI (kg/m^2)	25.0 (3.1)	25.4 (3.0)	25.5 (2.9)
Alcohol consumption (g ethanol/d)[c]	13.7 (15.0)	15.4 (15.3)	14.9 (14.7)
Energy intake (kcal)	1915 (505)	1998 (529)	1994 (517)
Diuretic medication (%)	11.0	14.0	15.2
Antihypertensive medication (%)	20.5	24.6	26.2

The subcohort includes 15 RCC cases, of which 11 ccRCC cases. Solely participants with complete information for main exposures are included in this table. SD = Standard deviation, RCC = renal cell carcinoma, ccRCC = clear cell renal cell carcinoma, BMI = Body Mass Index.
[a] According to the TNM version used at time of diagnosis by the Netherlands Cancer registry.
[b] In former and current smokers only.
[c] In alcohol consumers only.

Table 2 - Association between *VHL* and *HIF1A* Single Nucleotide Polymorphisms (SNPs) and (clear cell) renal cell carcinoma status; Netherlands Cohort Study on diet and cancer, 1986-2006

			Renal cell carcinoma				Clear-cell renal cell carcinoma			
SNP	HWE[a]	Subcohort person-years	No. cases	HR[b, c] (CI 95%)	$HR_{adjusted}$[c, d] (CI 95%)	*p* for trend[e]	No. cases	HR[b, c] (CI 95%)	$HR_{adjusted}$[c, d] (CI 95%)	*p* for trend[e]
*VHL*_rs1642739[f]	0.291									
GG		39067	299	1 (reference)	1 (reference)		190	1 (reference)	1 (reference)	
GT		11262	103	1.19 (0.93-1.52)	1.18 (0.92-1.51)		71	1.29 (0.97-1.72)	1.29 (0.96-1.73)	
TT		993	4	0.50 (0.18-1.40)	0.50 (0.18-1.41)		2	0.39 (0.10-1.64)	0.40 (0.10-1.66)	
T				1.06 (0.86-1.29)	1.05 (0.86-1.29)	0.635		1.11 (0.87-1.41)	1.11 (0.87-1.41)	0.410
GT+TT		12255	107	1.13 (0.89-1.44)	1.13 (0.89-1.43)		73	1.22 (0.91-1.62)	1.21 (0.91-1.62)	
*VHL*_rs779805	0.598									
AA		23908	157	1 (reference)	1 (reference)		98	1 (reference)	1 (reference)	
AG		22432	198	1.33 (1.06-1.66)	1.32 (1.06-1.66)		126	1.36 (1.03-1.79)	1.35 (1.02-1.78)	
GG		4982	51	1.50 (1.06-2.11)	1.53 (1.07-2.17)		39	1.83 (1.23-2.73)	1.88 (1.25-2.81)	
G				1.25 (1.07-1.46)	1.26 (1.08-1.47)	0.004		1.35 (1.12-1.63)	1.36 (1.13-1.65)	0.001
AG+GG		27414	249	1.36 (1.10-1.69)	1.36 (1.10-1.69)		165	1.45 (1.11-1.88)	1.44 (1.11-1.88)	
*VHL*_rs265318[f]	0.965									
AA		41066	321	1 (reference)	1 (reference)		206	1 (reference)	1 (reference)	
AC		9686	82	1.10 (0.85-1.43)	1.09 (0.83-1.42)		54	1.13 (0.82-1.55)	1.12 (0.81-1.54)	
CC		570	3	0.62 (0.19-2.07)	0.65 (0.19-2.20)		3	0.97 (0.29-3.23)	1.03 (0.30-3.54)	
C				1.03 (0.82-1.30)	1.03 (0.81-1.30)	0.809		1.09 (0.83-1.45)	1.09 (0.82-1.45)	0.534
AC+CC		10256	85	1.07 (0.83-1.38)	1.06 (0.82-1.38)		57	1.12 (0.82-1.52)	1.11 (0.81-1.52)	
*HIF1A*_rs2301111[f]	0.581									
CC		32553	257	1 (reference)	1 (reference)		170	1 (reference)	1 (reference)	
CG		16741	138	1.06 (0.85-1.33)	1.06 (0.84-1.32)		88	1.02 (0.78-1.34)	1.02 (0.78-1.34)	
GG		2028	11	0.76 (0.40-1.44)	0.77 (0.41-1.47)		5	0.52 (0.21-1.29)	0.52 (0.21-1.30)	
G				0.99 (0.83-1.19)	0.99 (0.83-1.20)	0.956		0.93 (0.74-1.16)	0.92 (0.74-1.16)	0.486
CG+GG		18769	149	1.03 (0.83-1.28)	1.03 (0.83-1.28)		93	0.97 (0.75-1.27)	0.97 (0.74-1.27)	

Reference/minor alleles for the selected SNPs in the NLCS: *VHL*_rs1642739 – G/T; *VHL*_rs779805 – A/G; *VHL*_rs265318 – A/C; *HIF1A*_rs2301111 – C/G

[a] Hardy-Weinberg Equilibrium as tested with the Pearson χ^2-test based on the distribution of genotypes in the subcohort as presented in Supplementary Table 1.

[b] Models adjusted for age (y, continuous) and sex (man/woman)

[c] Models include a time-varying covariable for age due to a violation of the proportional hazards assumption.

[d] Models adjusted for age (y, continuous), sex (man/woman), hypertension (yes/no), smoking status (never, former, current), smoking duration (y, centered), smoking intensity (cig/d, centered), BMI (kg/m^2, continuous) and alcohol consumption (g ethanol/d, continuous).

[e] The per-allele *p* for trend is based on the multivariable-adjusted model

[f] Minor allele frequency < 0.25

Gene-environment interactions

In multivariable-adjusted models for RCC risk, potential gene-environment interactions were observed between *VHL*_rs1642739*, VHL*_rs779805 and *HIF1A*_rs2301111 SNPs and alcohol consumption (Table 3). A weak inverse association between alcohol consumption (per 5g/day) and RCC risk was observed in participants carrying the rare genotype for *VHL*_rs1642739 and *VHL*_rs779805, but not in participants carrying the wild-type genotype. For carriers of the wild-type *HIF1A*_rs2301111 genotype a weak inverse association between alcohol consumption and RCC risk, but not for individuals carrying the rare genotype. No interaction was observed between either of the selected SNPs and self-reported hypertension (yes, no), smoking status (never, former, current) and BMI (per kg/m^2) for RCC risk. For ccRCC, a potential interaction between *VHL*_rs779805 SNPs and alcohol consumption was observed. However, after correction for multiple comparisons using the adaptive Benjamini-Hochberg method none of the potential gene-environment interactions maintained statistical significance[45].

In sensitivity analyses, a potential gene-environment interaction was apparent between categorized alcohol consumption (0 g/d, 0.1-4 g/d, 5-14 g/d, 15-29 g/d and 30+ g/d) and *VHL*_rs164273 status for ccRCC risk (p=0.009; Supplementary Table 2). The direction of associations for *VHL*_rs779805 was similar to main analyses using alcohol consumption (per 5g/day). Sensitivity analyses between smoking status (ever/never), hypertension (no self-reported hypertension or no self-reported antihypertensive medication, hypertension with self-reported hypertensive medication) and BMI (<20 kg/m^2, 20-<25 kg/m^2, 25-<30 kg/m^2 and 30+ kg/m^2) and SNP status showed similar associations compared to main gene-environment analyses (Supplementary Table 2). Similar to main analyses, no sensitivity analysis remained statistically significant after multiple comparison correction.

Gene-gene interactions

No gene-gene interactions, as tested with the Wald χ^2 test, were found between the three selected *VHL* SNPs and *HIF1A*_rs2301111 for both RCC (p=0.310, p=0.321 and p=0.514 for *VHL*_rs1642739*, VHL*_rs779805 and *VHL*_rs265318, respectively) and ccRCC (p=0.762, p=0.442 and p=0.978 for *VHL*_rs1642739*, VHL*_rs779805 and *VHL*_rs265318*,* respectively).

Association between SNPs and VHL promoter methylation status

In total, information on VHL promoter methylation was available from 253 ccRCC cases. Among ccRCC cases, 19 (7.5%) participants had a methylated CpG island in the *VHL* promoter region of which 13 had at least one mutant allele for the selected *VHL* SNPs (Supplementary Table 3). *VHL* promoter methylation was apparent in three, twelve and two participants for the rare genotype of *VHL*_rs1642739 (GG vs. GT+TT), *VHL*_rs779805 (AA vs. AG+GG) and *VHL*_rs264318 (AA vs. AC+CC), respectively. In multivariable-adjusted analyses a non-significant inverse association was observed between both *VHL*_rs1642739 (HR 0.45, 95%CI 0.12-1.69) and *VHL*_rs265318 (HR 0.38, 95%CI 0.07-2.00) and *VHL* promoter methylation in ccRCC cases. No association was observed for the *VHL*_rs779805 SNP (HR 0.99, 95%CI 0.37-2.69).

Table 3 – Multivariable-adjusted gene-environment interactions for renal cell carcinoma (RCC) and clear-cell renal cell carcinoma; Netherlands Cohort Study on diet and cancer, 1986-2006

Gene		Wild genotype			Rare genotype				
Environment		Subcohort person-years	No. cases	$HR_{adjusted}$ [a,b] (CI95)	Subcohort person-years	No. cases	$HR_{adjusted}$ [a,b] (CI95)	*P* for interaction[c]	FDR *p* threshold
Renal cell carcinoma									
*VHL*_rs1642739		GG			GT + TT				
Hypertension	No	28949	198	1 (reference)	9088	72	1 (reference)		
	Yes	10118	101	1.48 (1.13-1.92)	3167	35	1.48 (0.95-2.32)	0.968	0.100
Smoking status[d]	Never	15367	72	1 (reference)	4667	37	1 (reference)		
	Former	13829	135	1.41 (0.97-2.04)	4558	43	1.26 (0.74-2.14)		
	Current	9871	92	1.41 (0.97-2.05)	3029	27	1.13 (0.64-1.99)	0.136	0.031
BMI	Per kg/m^2	39067	299	1.05 (1.01-1.10)	12255	107	1.00 (0.94-1.06)	0.212	0.044
Alcohol consumption	Per 5g/day	39067	299	1.00 (0.96-1.05)	12255	107	0.91 (0.82-1.02)	0.023	0.006
*VHL*_rs779805		AA			AG + GG				
Hypertension	No	17764	109	1 (reference)	20273	161	1 (reference)		
	Yes	6145	48	1.34 (0.92-1.93)	7140	88	1.55 (1.16-2.08)	0.433	0.063
Smoking status[d]	Never	9542	37	1 (reference)	10492	72	1 (reference)		
	Former	8471	72	1.31 (0.80-2.13)	9916	106	1.37 (0.93-2.02)		
	Current	5895	48	1.38 (0.82-2.31)	7005	71	1.29 (0.87-1.91)	0.330	0.050
BMI	Per kg/m^2	23908	157	1.04 (0.98-1.11)	27414	249	1.04 (0.99-1.08)	0.833	0.081
Alcohol consumption	Per 5g/day	23908	157	1.02 (0.97-1.06)	27414	249	0.95 (0.90-1.01)	0.024	0.013
*VHL*_rs265318		AA			AC + CC				
Hypertension	No	30540	211	1 (reference)	7497	59	1 (reference)		
	Yes	10526	110	1.52 (1.18-1.96)	2759	26	1.32 (0.79-2.20)	0.553	0.069
Smoking status[d]	Never	16103	85	1 (reference)	3932	24	1 (reference)		
	Former	14813	141	1.32 (0.94-1.86)	3574	37	1.38 (0.74-2.60)		
	Current	10150	95	1.28 (0.90-1.83)	2750	24	1.43 (0.76-2.71)	0.909	0.088
BMI	Per kg/m^2	41066	321	1.05 (1.01-1.09)	10256	85	0.99 (0.92-1.07)	0.131	0.025
Alcohol consumption	Per 5g/day	41066	321	0.99 (0.95-1.03)	10256	85	0.95 (0.85-1.06)	0.359	0.056
*HIF1A*_rs2301111		CC			CG + GG				
Hypertension	No	24090	170	1 (reference)	13947	100	1 (reference)		
	Yes	8462	87	1.48 (1.12-1.97)	4822	49	1.53 (1.04-2.24)	0.934	0.094
Smoking status[d]	Never	12704	77	1 (reference)	7330	32	1 (reference)		
	Former	11620	108	1.17 (0.80-1.70)	6767	70	1.77 (1.07-2.92)		
	Current	8229	72	1.05 (0.71-1.54)	4672	47	2.01 (1.19-3.40)	0.167	0.038
BMI	Per kg/m^2	32553	257	1.04 (0.99-1.08)	18769	149	1.05 (0.99-1.11)	0.823	0.075
Alcohol consumption	Per 5g/day	32553	257	0.96 (0.91-1.00)	18769	149	1.04 (0.97-1.11)	0.024	0.019

Continues next page

Table 3 - continued

Clear-cell renal cell carcinoma									
*VHL*_rs1642739		GG			GT + TT				
Hypertension	No	28949	124	1 (reference)	9088	51	1 (reference)		
	Yes	10118	66	1.51 (1.09-1.09)	3167	22	1.38 (0.80-2.38)	0.614	0.069
Smoking status[d]	Never	15367	47	1 (reference)	4667	23	1 (reference)		
	Former	13829	90	1.56 (1.01-2.41)	4558	29	1.36 (0.73-2.56)		
	Current	9871	53	1.33 (0.83-2.12)	3029	21	1.34 (0.69-2.59)	0.398	0.050
BMI	Per kg/m^2	39067	190	1.06 (1.01-1.11)	12255	73	1.00 (0.93-1.07)	0.103	0.019
Alcohol intake	Per 5g/day	39067	190	0.99 (0.95-1.05)	12255	73	0.90 (0.77-1.04)	0.085	0.012
*VHL*_rs779805		AA			AG + GG				
Hypertension	No	17764	66	1 (reference)	20273	109	1 (reference)		
	Yes	6145	32	1.45 (0.92-2.28)	7140	56	1.47 (1.03-2.09)	0.929	0.094
Smoking status[d]	Never	9542	21	1 (reference)	10492	49	1 (reference)		
	Former	8471	49	1.65 (0.89-3.06)	9916	70	1.42 (0.91-2.22)		
	Current	5895	28	1.64 (0.84-3.21)	7005	46	1.19 (0.74-1.91)	0.200	0.038
BMI	Per kg/m^2	23908	98	1.06 (0.98-1.13)	27414	165	1.04 (0.99-1.09)	0.853	0.088
Alcohol intake	Per 5g/day	23908	98	1.02 (0.96-1.08)	27414	165	0.94 (0.87-1.01)	0.020	0.006
*VHL*_rs265318		AA			AC + CC				
Hypertension	No	30540	135	1 (reference)	7497	40	1 (reference)		
	Yes	10526	71	1.53 (1.12-2.08)	2759	17	1.25 (0.68-2.31)	0.544	0.056
Smoking status[d]	Never	16103	54	1 (reference)	3932	16	1 (reference)		
	Former	14813	95	1.47 (0.98-2.22)	3574	24	1.46 (0.70-303)		
	Current	10150	57	1.24 (0.80-1.94)	2750	17	1.59 (0.76-3.33)	0.962	0.100
BMI	Per kg/m^2	41066	206	1.05 (1.00-1.10)	10256	57	1.01 (0.94-1.09)	0.191	0.031
Alcohol consumption	Per 5g/day	41066	206	0.99 (0.94-1.04)	10256	57	0.90 (0.76-1.06)	0.161	0.025
*HIF1A*_rs2301111		CC			CG + GG				
Hypertension	No	24090	114	1 (reference)	13947	61	1 (reference)		
	Yes	8462	56	1.43 (1.02-2.01)	4822	32	1.61 (1.01-2.57)	0.640	0.075
Smoking status[d]	Never	12704	48	1 (reference)	7330	22	1 (reference)		
	Former	11620	74	1.34 (0.86-2.08)	6767	45	1.80 (0.98-3.32)		
	Current	8229	48	1.18 (0.74-1.88)	4672	26	1.68 (0.86-3.27)	0.657	0.081
BMI	Per kg/m^2	32553	170	1.03 (0.98-1.08)	18769	93	1.06 (0.99-1.13)	0.545	0.063
Alcohol consumption	Per 5g/day	32553	170	0.96 (0.90-1.02)	18769	93	1.01 (0.93-1.10)	0.319	0.044

[a] Models include a time-varying covariable for age due to a probable violation of the proportional hazards assumption.
[b] Models adjusted for age (y, continuous), sex (man/woman), hypertension (yes/no), smoking status (never, former, current), smoking duration (y, centered), smoking intensity (cig/d, centered), BMI (kg/m^2, continuous) and alcohol consumption (g ethanol/d, continuous) when applicable
[c] No interaction was significant after false discovery rate correction
[d] Additionally adjusted for smoking duration (centered, years) and smoking intensity (centered, cig/d)

Discussion

In this study, a statistically significantly increased RCC risk was found for individuals that carry genotypes with at least one variant allele for the *VHL*_rs779805 SNP. This association was especially pronounced for ccRCC risk. No association was found for *VHL*_rs164239, *VHL*_rs265318 and *HIF1A*_rs2301111. After adjustment for multiple comparisons, no statistically significant gene-environment interactions were found between the selected SNPs and smoking, hypertension, BMI and alcohol for both RCC and ccRCC cases. No gene-gene interactions were found between selected *VHL* SNPs and the *HIF1A* SNP.

Several studies have assessed the relationship between the *VHL*_rs779805 SNP and sporadic RCC[19-21]. Lv *et al.* found an association between the germline SNP *VHL*_rs779805 and RCC risk. Similarly, we found a statistically significant positive trend for the G allele and a positive association between the GG genotype for *VHL*_rs779805 and RCC risk[20]. The aforementioned studies did not report associations between *VHL* SNPs and ccRCC risk. In our study, rare *VHL*_rs779805 genotypes had a stronger association with ccRCC risk than with RCC risk. This might indicate that *VHL* polymorphisms lead to an increased susceptibility for ccRCC in particular. To our knowledge, no other study has investigated the relationship between *VHL*_rs1642739, *VHL*_rs265318 and *HIF1A*_rs2301111 and (cc)RCC risk. In this study, no association was found between (cc)RCC risk and *VHL*_rs1642739, *VHL*_rs265318 or *HIF1A*_rs2301111.

Multiple studies have assessed gene-environment interactions in RCC and ccRCC. RCC risk has been found to be associated with interactions between alcohol consumption and *ADH7*[26]; sodium and hypertension and *AGTR*, *AGT* and *ACE*[34]; calcium and vitamin D intake and *RXRA*[28]; tobacco smoking and *NAT2*, *CYP1A1* and *GSTM1*[25]; and meat-cooking mutagens and *ITPR2* and *EPAS1*[27]. To our knowledge, we are the first to study gene-environment interactions between the selected *VHL* and *HIF1A* SNPs and smoking, hypertension, BMI and alcohol consumption. Solely the interaction between *VHL*_rs779805 and alcohol consumption was associated with both RCC and ccRCC risk. However, this association did not maintain statistical significance after correction for multiple comparisons with the adaptive Benjamini-Hochberg method. Dominant models were used for all gene-environment analyses because of the low MAF of most included SNPs. However, SNPs may not have adhered to a dominant model, as there may be differences in disease susceptibility between heterozygous and homozygous rare genotypes, as was found for *VHL*_rs779805 (Table 2). This exemplifies that our gene-environment analyses may have been hampered by the inability to assess interactions per genotype. Further research is needed to ascertain the interaction between alcohol and *VHL* SNP status on (cc)RCC risk.

Disruptions in the *VHL* tumor suppressor gene are thought to play a role in the constitutive activation of hypoxia-inducible factors, as regulated in part by *HIF1A*, which may lead to carcinogenesis[1]. Therefore, it is plausible for gene-gene interactions to occur. However, in this study, we did not find gene-gene interactions between selected *VHL* and *HIF1A* SNPs on the risk of developing (cc)RCC.

Previous studies have found a relationship between VHL promoter hypermethylation and SNPs in *VHL*_rs779805 in sporadic ccRCC cases[6, 23]. Moore *et al.* also reported a positive

association between promoter hypermethylation and *VHL*_rs265318 and *VHL*_rs1642739. In contrast, we found no association between promoter methylation status and *VHL*_rs779805 in ccRCC cases. *VHL*_rs1642739 and *VHL*_rs265318 seemed inversely associated with *VHL* promoter methylation in ccRCC cases. However, this association was based on a limited sample size. While the number of cases with known promoter methylation status was similar in size to the study of Moore *et al.*, our study had a smaller proportion of cases with *VHL* promoter methylation (7.5% vs. 9.8%)[23]. Banks *et al.* reported an even higher proportion of sporadic ccRCC cases with a methylated *VHL* promoter (20.4%), but had a smaller study population[6]. In general, there are large differences in the proportion of methylated *VHL* promoters per SNP between studies, which may explain these unstable point estimates [23]. Therefore, more research with a larger number of sporadic ccRCC cases is needed to elucidate the relationship between *VHL* promoter methylation and *VHL* SNPs.

At present, genome-wide association studies (GWAS) have identified multiple novel risk loci that may contribute to RCC susceptibility. Interestingly, SNPs in the *VHL* and *HIF1A* genes have not (yet) been identified as potential risk variants, while there is a biological plausibility for the involvement of these genes based on current evidence on the development of RCC[2, 9]. For example, risk loci have been identified in *EPAS1*[11, 13, 17, 18], which is known to be involved in the VHL-HIF-1 pathway[47]. While we found no evidence for an association between three of our selected SNPs, *VHL*_rs779805 was associated with an increased risk of RCC. This finding was in line with two prior published studies, in which a potential association between *VHL*_rs779805 and RCC risk was found[19, 20]. While this particular SNP is present on commonly used SNP arrays, this SNP remains unidentified in large-scale GWAS studies[11-18]. It is estimated that the currently available risk loci for RCC account for approximately 10% of the familial risk for RCC[11]. Therefore, it may well be possible for minor susceptibility loci to remain unidentified in GWAS studies, due to their tendency to convey small-to-moderate changes in risk, while major susceptibility loci are detectable in the stringent false discovery rate correction criteria of GWAS studies. This could be a reason why SNPs like *VHL*_rs779805 may remain unidentified, unless alternative methodologies are employed[11]. As a result, there is ample opportunity to discover new, rarer, RCC risk variants in future research. Additional evidence on risk loci from GWAS studies, combined with extensive information on direct effects, environmental factors and other potential modulators of disease etiology from candidate SNP studies, should lead to new insights into the biology of RCC to further the potential for new prevention, early detection and intervention strategies to be employed[11].

This study also has several strengths. Strengths of this study were the detailed questionnaire information, the long duration and the histological revision of RCC cases by two experienced pathologists. Furthermore, cases in our study were obtained prospectively from a population of 120,852 men and women from 204 Dutch municipalities. Combined with the completeness of follow-up, we assume that these cases are a representative of kidney cancer cases in the Netherlands at the time.

In conclusion, this study confirmed the association between germline SNP *VHL*_rs779805 with RCC risk. In addition, a slightly stronger association for ccRCC was found compared to RCC. Potential gene-environment interactions were found between alcohol and *VHL* SNPs. However, results did not remain statistically significant after correction for multiple

comparisons. No gene-gene interactions were observed between the *VHL* and *HIF1A* SNPs. Lastly, tumor promoter methylation was not significantly associated with *VHL* SNPs.

References

1. Hsieh JJ, Purdue MP, Signoretti S, et al. Renal cell carcinoma. Nature reviews. Disease primers. 2017;3: 17009.
2. Kaelin WG. Von Hippel-Lindau disease. Annual review of pathology. 2007;2: 145-173.
3. Kim WY, Kaelin WG. Role of VHL Gene Mutation in Human Cancer. Journal of Clinical Oncology. 2004;22: 4991-5004.
4. Nickerson ML, Jaeger E, Shi Y, et al. Improved identification of von Hippel-Lindau gene alterations in clear cell renal tumors. Clinical cancer research : an official journal of the American Association for Cancer Research. 2008;14: 4726-4734.
5. van Houwelingen KP, van Dijk BA, Hulsbergen-van de Kaa CA, et al. Prevalence of von Hippel-Lindau gene mutations in sporadic renal cell carcinoma: results from The Netherlands cohort study. BMC Cancer. 2005;5: 57.
6. Banks RE, Tirukonda P, Taylor C, et al. Genetic and epigenetic analysis of von Hippel-Lindau (VHL) gene alterations and relationship with clinical variables in sporadic renal cancer. Cancer research. 2006;66: 2000-2011.
7. Gnarra JR, Tory K, Weng Y, et al. Mutations of the VHL tumour suppressor gene in renal carcinoma. Nature genetics. 1994;7: 85-90.
8. Cancer Genome Atlas Research N. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013;499: 43-49.
9. Schodel J, Grampp S, Maher ER, et al. Hypoxia, Hypoxia-inducible Transcription Factors, and Renal Cancer. Eur. Urol. 2016;69: 646-657.
10. Hunter DJ. Gene-environment interactions in human diseases. Nature reviews. Genetics. 2005;6: 287-298.
11. Scelo G, Purdue MP, Brown KM, et al. Genome-wide association study identifies multiple risk loci for renal cell carcinoma. Nature communications. 2017;8: 15724.
12. Henrion M, Frampton M, Scelo G, et al. Common variation at 2q22.3 (ZEB2) influences the risk of renal cancer. Human molecular genetics. 2013;22: 825-831.
13. Purdue MP, Johansson M, Zelenika D, et al. Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. Nature genetics. 2011;43: 60-65.
14. Wu X, Scelo G, Purdue MP, et al. A genome-wide association study identifies a novel susceptibility locus for renal cell carcinoma on 12p11.23. Human molecular genetics. 2012;21: 456-462.
15. Gudmundsson J, Sulem P, Gudbjartsson DF, et al. A common variant at 8q24.21 is associated with renal cell cancer. Nature communications. 2013;4: 2776.
16. Purdue MP, Ye Y, Wang Z, et al. A genome-wide association study of renal cell carcinoma among African Americans. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2014;23: 209-214.
17. Henrion MY, Purdue MP, Scelo G, et al. Common variation at 1q24.1 (ALDH9A1) is a potential risk factor for renal cancer. PloS one. 2015;10: e0122589.
18. Laskar RS, Muller DC, Li P, et al. Sex specific associations in genome wide association analysis of renal cell carcinoma. European journal of human genetics : EJHG. 2019;27: 1589-1598.
19. Wang WC, Tsou MH, Chen HJ, Hsu WF, Lai YC. Two single nucleotide polymorphisms in the von Hippel-Lindau tumor suppressor gene in Taiwanese with renal cell carcinoma. BMC research notes. 2014;7: 638.

20. Lv C, Bai Z, Liu Z, Luo P, Zhang J. Renal cell carcinoma risk is associated with the interactions of APOE, VHL and MTHFR gene polymorphisms. International journal of clinical and experimental pathology. 2015;8: 5781-5786.
21. Qin C, Cao Q, Ju X, et al. The polymorphisms in the VHL and HIF1A genes are associated with the prognosis but not the development of renal cell carcinoma. Annals of oncology : official journal of the European Society for Medical Oncology. 2012;23: 981-989.
22. Qin C, Chen J, Li J, et al. Variants in angiogenesis-related genes and the risk of clear cell renal cell carcinoma. Mutagenesis. 2014;29: 419-425.
23. Moore LE, Nickerson ML, Brennan P, et al. Von Hippel-Lindau (VHL) inactivation in sporadic clear cell renal cancer: associations with germline VHL polymorphisms and etiologic risk factors. PLoS genetics. 2011;7: e1002312.
24. Schouten LJ, Deckers IA, van den Brandt PA, Baldewijns MM, van Engeland M. Alcohol and Dietary Folate Intake and Promoter CpG Island Methylation in Clear-Cell Renal Cell Cancer. Nutrition and cancer. 2016;68: 1097-1107.
25. Smits KM, Schouten LJ, van Dijk BA, et al. Polymorphisms in genes related to activation or detoxification of carcinogens might interact with smoking to increase renal cancer risk: results from The Netherlands Cohort Study on diet and cancer. World journal of urology. 2008;26: 103-110.
26. Antwi SO, Eckel-Passow JE, Diehl ND, et al. Alcohol consumption, variability in alcohol dehydrogenase genes and Risk of Renal Cell Carcinoma. International journal of cancer. 2017.
27. Melkonian SC, Daniel CR, Ye Y, et al. Gene-environment interaction of genome-wide association study-identified susceptibility loci and meat-cooking mutagens in the etiology of renal cell carcinoma. Cancer. 2016;122: 108-115.
28. Karami S, Brennan P, Navratilova M, et al. Vitamin d pathway genes, diet, and risk of renal cell carcinoma. International journal of endocrinology. 2010;2010: 879362.
29. Chow WH, Dong LM, Devesa SS. Epidemiology and risk factors for kidney cancer. Nature reviews. Urology. 2010;7: 245-257.
30. van den Brandt PA, Goldbohm RA, van ‘t Veer P, Volovics A, Hermus RJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. Journal of clinical epidemiology. 1990;43: 285-295.
31. Van den Brandt PA, Schouten LJ, Goldbohm RA, Dorant E, Hunen PM. Development of a record linkage protocol for use in the Dutch Cancer Registry for Epidemiological Research. Int. J. Epidemiol. 1990;19: 553-558.
32. Goldbohm RA, Van den Brandt PA, Dorant E. Estimation of the coverage of Dutch municipalities by cancer registries and PALGA based on hospital discharge data. Tijdschr Soc Gezondheidsz. 1994;72: 80-84.
33. Eble J, Sauter G, Epstein J, Sesterhenn I. World Health Organization Classification of Tumours. Pathology and Genetics. Tumours of the Urinary System and Male Genital Organs. Lyon: IARC Press, 2004.
34. Deckers IA, van den Brandt PA, van Engeland M, et al. Polymorphisms in genes of the renin-angiotensin-aldosterone system and renal cell cancer risk: interplay with hypertension and intakes of sodium, potassium and fluid. Int J Cancer. 2015;136: 1104-1116.
35. Hogervorst JG, Godschalk RW, van den Brandt PA, et al. DNA from nails for genetic analyses in large-scale epidemiologic studies. Cancer Epidemiol Biomarkers Prev. 2014;23: 2703-2712.
36. Cline RE, Laurent NM, Foran DR. The fingernails of Mary Sullivan: developing reliable methods for selectively isolating endogenous and exogenous DNA from evidence. J Forensic Sci. 2003;48:

328-333.
37. van Houwelingen KP, van Dijk BA, Hulsbergen-van de Kaa CA, et al. Prevalence of von Hippel-Lindau gene mutations in sporadic renal cell carcinoma: results from The Netherlands cohort study. BMC Cancer. 2005;5: 57.
38. Schouten LJ, van Dijk BA, Oosterwijk E, et al. Alcohol consumption and mutations or promoter hypermethylation of the von Hippel-Lindau gene in renal cell carcinoma. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2008;17: 3543-3550.
39. Derks S, Lentjes MH, Hellebrekers DM, de Bruine AP, Herman JG, van Engeland M. Methylation-specific PCR unraveled. Cellular oncology : the official journal of the International Society for Cellular Oncology. 2004;26: 291-299.
40. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proceedings of the National Academy of Sciences of the United States of America. 1996;93: 9821-9826.
41. van Engeland M, Roemen GM, Brink M, et al. K-ras mutations and RASSF1A promoter methylation in colorectal cancer. Oncogene. 2002;21: 3792-3795.
42. van den Brandt PA, Goldbohm RA. A prospective cohort study on diet and cancer in the Netherlands : design, conduct, analysis and first results after 3.3 years of follow-up. Datawyse / Universitaire Pers Maastricht. Maastricht: Maastricht University, 1993:208.
43. Goldbohm RA, Van den Brandt PA, Van ‘t Veer P, Dorant E, Sturmans F, Hermus RJ. Prospective study on alcohol consumption and the risk of cancer of the colon and rectum in the Netherlands. Cancer causes & control : CCC. 1994;5: 95-104.
44. Cleves MA. Exploratory analysis of single nucleotide polymorphism (SNP) for quantitative traits. Stata Journal. 2005;5: 141-153.
45. Benjamini Y, Hochberg Y. On the Adaptive Control of the False Discovery Rate in Multiple Testing with Independent Statistics. Journal of Educational and Behavioral Statistics. 2000;25: 60-83.
46. Schoenfeld D. Partial Residuals for The Proportional Hazards Regression Model. Biometrika. 1982;69: 239-241.
47. Schödel J, Bardella C, Sciesielski LK, et al. Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. Nature genetics. 2012;44: 420-S422.

Supplementary Table 1 - Descriptive characteristics of SNPs in subcohort members with ≥95% sample call rate; Netherlands Cohort Study on diet and cancer

					Genotype[a]							
					11		12		22		HWE[b]	
SNP	Gene	Chr. Location	Ref. Alleles	Minor Allele	n	%	n	%	n	%	p-value	MAF[c]
rs1642739	*VHL*	3p25.3	G/T	T	2283	76.00	664	22.10	57	1.90	0.291	0.13
rs779805	*VHL*	3p25.3	A/G	G	1414	47.07	1303	43.38	287	9.55	0.598	0.31
rs265318	*VHL*	3p25.3	A/C	C	2401	79.93	569	18.94	34	1.13	0.965	0.11
rs2301111	*HIF1A*	14q23.2	C/G	G	1895	63.08	988	32.89	121	4.03	0.581	0.20

[a] 11: homozygous for major allele, 12: heterozygous for major and minor alleles, 22: homozygous for minor allele
[b] Hardy-Weinberg Equilibrium as tested with Pearson χ^2 test
[c] Minor Allele Frequency

Supplementary Table 2 - Sensitivity analyses for multivariable-adjusted gene-environment interactions for renal cell carcinoma (RCC) and clear-cell renal cell carcinoma; Netherlands Cohort Study on diet and cancer, 1986-2006

Gene		Wild genotype			Rare genotype				
Environment		Subcohort Person-years	No. cases	$HR_{adjusted}$ [a,b] (CI95)	Subcohort Person-years	No. cases	$HR_{adjusted}$ [a,b] (CI95)	*P* for interaction[c]	FDR *p* threshold
Renal cell carcinoma									
*VHL*_rs1642739		GG			GT + TT				
Hypertension	No	33303	85	1	10653	10	1		
Incl. hypertensive meds	Yes	5764	22	1.71 (1.26-2.33)	1602	4	1.90 (1.09-3.29)	0.855	0.081
Smoking status[d]	Never	15367	72	1	4667	37	1		
	Ever	23700	227	1.41 (1.02-1.95)	7588	70	1.19 (0.75-1.88)	0.046	0.013
BMI	<20 kg/m^2	1210	4	0.61 (0.22-1.72)	422	2	0.62 (0.14-2.68)		
	20-<25	20729	138	1	5906	49	1		
	25-<30	14748	141	1.32 (1.02-1.71)	5024	50	1.08 (0.68-1.73)		
	30+	2380	16	1.02 (0.57-1.83)	903	6	0.74 (0.47-1.18)	0.858	0.088
Alcohol intake	0 g/d	9220	59	1	2630	29	1		
	0.1-4 g/d	11481	72	0.91 (0.63-1.31)	3732	33	0.75 (0.43-1.32)		
	5-14 g/d	8695	60	0.79 (0.54-1.17)	2646	24	0.68 (0.35-1.33)		
	15-29 g/d	6132	72	1.17 (0.79-1.74)	2083	13	0.41 (0.18-0.92)		
	>=30 g/d	3540	36	0.89 (0.55-1.44)	1164	8	0.42 (0.16-1.08)	0.019	0.006
*VHL*_rs779805		AA			AG + GG				
Hypertension	No	20435	127	1	23521	192	1		
Incl. hypertensive meds	Yes	3473	30	1.62 (1.04-2.52)	3892	57	1.79 (1.27-2.52)	0.529	0.056
Smoking status[d]	Never	9542	37	1	10492	72	1		
	Ever	14366	120	1.34 (0.86-2.08)	16922	177	1.33 (0.95-1.86)	0.137	0.031
BMI	<20 kg/m^2	820	3	0.87 (0.26-2.90)	811	3	0.51 (0.16-1.66)		
	20-<25	12555	71	1	14080	116	1		
	25-<30	8902	75	1.40 (0.98-2.00)	10870	116	1.18 (0.89-1.57)		
	30+	1631	8	1.07 (0.47-2.39)	1652	14	0.91 (0.49-1.69)	0.898	0.093
Alcohol intake	0 g/d	5691	25	1	6159	63	1		
	0.1-4 g/d	7250	39	1.08 (0.63-1.83)	7963	66	0.78 (0.54-1.14)		
	5-14 g/d	5035	34	1.12 (0.63-1.97)	6306	50	0.61 (0.40-0.93)		
	15-29 g/d	3796	39	1.40 (0.78-2.49)	4419	46	0.71 (045-1.11)		
	>=30 g/d	2137	20	1.09 (0.55-2.17)	2567	24	0.59 (0.34-1.03)	0.096	0.025
*VHL*_rs265318		AA			AC + CC				
Hypertension	No	35163	251	1	8793	68	1		
Incl. hypertensive meds	Yes	5903	70	1.77 (1.31-2.37)	1463	17	1.69 (0.90-3.16)	0.705	0.069
Smoking status[d]	Never	16103	85	1	3932	24	1		
	Ever	24964	236	1.30 (0.96-1.76)	6324	61	1.41 (0.82-2.42)	0.815	0.075

Continues on next page

Supplementary Table 2 - Continued

Gene		Wild genotype			Rare genotype				
Environment		Subcohort Person-years	No. cases	$HR_{adjusted}$ [a,b] (CI95)	Subcohort Person-years	No. cases	$HR_{adjusted}$ [a,b] (CI95)	*P* for interaction[c]	FDR *p* threshold
Renal cell carcinoma									
*VHL*_rs265318		AA			AC + CC				
BMI	<20 kg/m²	1294	5	0.71 (0.28-1.79)	337	1	0.40 (0.53-3.00)		
	20-<25	21654	143	1	4981	44	1		
	25-<30	15468	155	1.40 (1.09-1.79)	4304	36	0.84 (0.49-1.43)		
	30+	2650	18	1.03 (0.60-1.78)	634	4	0.78 (0.25-2.40)	0.307	0.044
Alcohol intake	0 g/d	9649	67	1	2201	21	1		
	0.1-4 g/d	12140	80	0.89 (0.63-1.26)	3073	25	0.78 (0.41-1.48)		
	5-14 g/d	9071	66	0.81 (0.56-1.17)	2270	18	0.64 (0.31-1.35)		
	15-29 g/d	6514	72	1.06 (0.72-1.55)	1701	13	0.54 (0.23-1.28)		
	>=30 g/d	3693	36	0.83 (0.52-1.34)	1011	8	0.50 (0.18-1.37)	0.503	0.050
*HIF1A*_rs2301111		CC			CG + GG				
Hypertension	No	27916	202	1	16040	117	1		
Incl. hypertensive meds	Yes	4636	55	1.71 (1.23-2.38)	2729	32	1.79 (1.13-2.83)	0.901	0.100
Smoking status[d]	Never	12704	77	1	7330	32	1		
	Ever	19849	180	1.11 (0.80-1.53)	11439	117	1.88 (1.19-2.96)	0.061	0.019
BMI	<20 kg/m²	1085	4	0.66 (0.24-1.85)	547	2	0.59 (0.14-2.51)		
	20-<25	16712	115	1	9923	72	1		
	25-<30	12538	126	1.35 (1.02-1.79)	7234	65	1.17 (0.81-1.70)		
	30+	218	12	0.80 (0.41-1.53)	1065	10	1.36 (0.64-2.88)	0.569	0.063
Alcohol intake	0 g/d	7176	54	1	4674	34	1		
	0.1-4 g/d	9527	72	0.92 (0.63-1.36)	5685	33	0.75 (0.45-1.26)		
	5-14 g/d	7321	58	0.79 (0.52-1.20)	4020	26	0.72 (0.41-1.25)		
Hypertension	15-29 g/d	5264	46	0.77 (0.49-1.23)	2951	39	1.16 (0.67-2.01)		
	>=30 g/d	3265	27	0.65 (0.38-1.11)	1438	17	0.97 (0.48-1.95)	0.224	0.038
Clear cell Renal Cell Carcinoma									
*VHL*_rs1642739		GG			GT + TT				
Hypertension	No	33303	148	1	10653	59	1		
Incl. hypertensive meds	Yes	5764	42	1.71 (1.18-2.48)	1602	14	1.88 (0.96-3.69)	0.930	0.088
Smoking status[d]	Never	15367	47	1	4667	23	1		
	Ever	23700	143	1.45 (0.98-2.15)	7588	50	1.35 (0.79-2.32)	0.287	0.031
BMI	<20 kg/m²	1210	3	0.76 (0.23-2.49)	422	1	0.46 (0.06-3.47)		
	20-<25	20729	83	1	5906	32	1		
	25-<30	14748	93	1.45 (1.06-1.98)	5024	38	1.34 (0.80-2.24)		
	30+	2380	11	1.14 (0.56-2.29)	903	2	0.43 (0.10-1.89)	0.720	0.069

Continues on next page

Supplementary Table 2 - Continued

Gene		Wild genotype			Rare genotype				
Environment		Subcohort Person-years	No. cases	$HR_{adjusted}$ [a,b] (CI95)	Subcohort Person-years	No. cases	$HR_{adjusted}$ [a,b] (CI95)	*P* for interaction[c]	FDR *p* threshold
Clear cell Renal Cell Carcinoma									
*VHL*_rs1642739		GG			GT + TT				
Alcohol intake	0 g/d	9220	40	1	2630	20	1		
	0.1-4 g/d	11481	40	0.74 (0.47-1.17)	3732	26	0.81 (0.43-1.55)		
	5-14 g/d	8695	40	0.80 (0.51-1.26)	2646	14	0.51 (0.22-1.18)		
	15-29 g/d	6132	51	1.29 (0.81-2.05)	2083	6	0.24 (0.08-0.67)		
	>=30 g/d	3540	19	0.74 (0.40-1.37)	1164	7	0.48 (0.17-1.38)	0.009	0.006
*VHL*_rs779805		AA			AG + GG				
Hypertension	No	20435	79	1	23521	128	1		
Incl. hypertensive meds	Yes	3473	19	1.76 (1.17-2.64)	3892	37	1.41 (1.05-1.89)	0.741	0.075
Smoking status[d]	Never	9542	21	1	10492	49	1		
	Ever	14366	77	1.65 (0.93-2.90)	16922	116	1.31 (0.88-1.93)	0.084	0.019
BMI	<20 kg/m^2	820	3	1.50 (0.44-5.12)	811	1	0.26 (0.04-1.93)		
	20-<25	12555	41	1	14080	74	1		
	25-<30	8902	48	1.56 (1.00-2.42)	10870	83	1.35 (0.97-1.89)		
	30+	1631	6	1.33 (0.51-3.49)	1652	7	0.75 (0.32-1.74)	0.518	0.063
Alcohol intake	0 g/d	5691	15	1	6159	45	1		
	0.1-4 g/d	7250	22	1.01 (0.51-1.98)	7963	44	0.71 (0.46-1.10)		
	5-14 g/d	5035	25	1.36 (0.69-2.68)	6306	29	0.48 (0.29-0.81)		
	15-29 g/d	3796	24	1.45 (0.70-2.68)	4419	33	0.71 (0.43-1.20)		
	>=30 g/d	2137	12	1.15 (0.49-2.73)	2567	14	0.49 (0.25-0.97)	0.049	0.013
*VHL*_rs265318		AA			AC + CC				
Hypertension	No	35163	162	1	8793	45	1		
Incl. hypertensive meds	Yes	5903	44	1.71 (1.19-2.45)	1463	12	1.86 (0.90-3.85)	0.955	0.094
Smoking status[d]	Never	16103	54	1	3932	16	1		
	Ever	24964	152	1.36 (0.94-1.97)	6324	41	1.53 (0.81-2.88)	0.847	0.081
BMI	<20 kg/m^2	1294	4	0.94 (0.33-2.63)	337	-	-		
	20-<25	21654	88	1	4981	27	1		
	25-<30	15468	103	1.51 (1.12-2.04)	4304	28	1.17 (0.66-2.08)		
	30+	2650	11	1.01 (0.51-2.02)	634	2	0.69 (0.15-3.19)	*	0.100
Alcohol intake	0 g/d	9649	44	1	2201	16	1		
	0.1-4 g/d	12140	47	0.80 (0.52-1.22)	3073	19	0.75 (0.36-1.55)		
	5-14 g/d	9071	43	0.80 (0.52-1.25)	2270	11	0.48 (0.20-1.17)		
	15-29 g/d	6514	51	1.16 (0.74-1.83)	1701	6	0.31 (0.10-0.91)		
	>=30 g/d	3693	21	0.76 (0.42-1.37)	1011	5	0.42 (0.13-1.42)	0.120	0.025

Continues on next page

Supplementary Table 2 - Continued

Gene		Wild genotype			Rare genotype				
Environment		Subcohort Person-years	No. cases	$HR_{adjusted}$ [a,b] (CI95)	Subcohort Person-years	No. cases	$HR_{adjusted}$ [a,b] (CI95)	*P* for interaction[c]	FDR *p* threshold
Clear cell Renal Cell Carcinoma									
*HIF1A*_rs2301111		CC			CG + GG				
Hypertension	No	27916	137	1	16040	70	1		
Incl. hypertensive meds	Yes	4636	33	1.53 (1.02-2.29)	2729	23	2.12 (1.23-3.65)	0.290	0.038
Smoking status[d]	Never	12704	48	1	7330	22	1		
	Ever	19849	122	1.26 (0.86-1.85)	11439	71	1.74 (1.00-3.05)	0.366	0.040
BMI	<20 kg/m^2	1085	3	0.76 (0.24-2.48)	547	1	0.51 (0.07-3.92)		
	20-<25	16712	74	1	9923	41	1		
	25-<30	12538	87	1.47 (1.05-2.05)	7234	44	1.37 (0.87-2.14)		
	30+	218	6	0.61 (0.25-1.48)	1065	7	1.64 (0.67-4.05)	0.411	0.056
Alcohol intake	0 g/d	7176	36	1	4674	24	1		
	0.1-4 g/d	9527	45	0.86 (0.54-1.36)	5685	21	0.67 (0.36-1.23)		
	5-14 g/d	7321	40	0.80 (0.49-1.31)	4020	14	0.53 (0.26-1.08)		
	15-29 g/d	5264	31	0.78 (0.45-1.35)	2951	26	1.13 (0.58-2.18)		
	>=30 g/d	3265	18	0.66 (0.35-1.25)	1438	8	0.70 (0.28-1.77)	0.386	0.050

[a] Models include a time-varying covariable for age due to a probable violation of the proportional hazards assumption.
[b] Models adjusted for age (y, continuous), sex (man/woman), hypertension (yes/no), smoking status (never, former, current), smoking duration (y, centered), smoking intensity (cig/d, centered), BMI (kg/m^2, continuous) and alcohol consumption (g ethanol/d, continuous) when applicable
[c] No interaction was significant after false discovery rate correction
[d] Additionally adjusted for smoking duration (centered, years) and smoking intensity (centered, cig/d)

Supplementary Table 3 - Case-only analysis on the association between *VHL* promoter methylation status and *VHL* SNP status, the Netherlands Cohort Study on Diet and Cancer, 1986-2006.

	VHL promoter methylation status		Age and sex-adjusted	Multivariable-adjusted
	Methylated	Unmethylated	OR[a] (95% CI)	OR[b] (95% CI)
Clear-cell renal cell carcinoma				
*VHL*_rs1642739				
GG	16	169	1	1
GT+TT	3	65	0.48 (0.14-1.73)	0.45 (0.12-1.69)
*VHL*_rs779805				
AA	7	88	1	1
AG+GG	12	146	1.02 (0.38-2.70)	0.99 (0.37-2.69)
*VHL*_rs265318				
AA	17	183	1	1
AC+CC	2	51	0.42 (0.09-1.87)	0.38 (0.07-2.00)
Total number of SNPs present				
None	6	85	1	1
1 or more	13	150	1.20 (0.43-3.36)	1.13 (0.39-3.29)

[a] Models adjusted for age (y, continuous) and sex (man/woman)
[b] Models adjusted for age (y, continuous), sex (man/woman), hypertension (yes/no), smoking status (never, former, current), smoking duration (y, centered), smoking intensity (cig/d, centered), BMI (kg/m^2, continuous) and alcohol intake (g ethanol/d, continuous)

CHAPTER 6

Evaluation of a seven-gene mutational profile as a prognostic factor in a population-based study of clear cell renal cell carcinoma

Jeroen AA van de Pol*, Paranita Ferronika*, Helga Westers, Manon van Engeland, Martijn M Terpstra, Kim Smits, Kim de Lange, Piet A van den Brandt, Rolf H Sijmons, Leo J Schouten and Klaas Kok

*Both authors contributed equally as first authors

Submitted for publication

CHAPTER 7

General discussion

In this chapter, the main findings of the chapters in this thesis will be discussed. Subsequently, we will critically reflect on methodological considerations and the strengths and weaknesses that may have impacted the findings in this thesis. Furthermore, we will discuss the impact of the main findings. Moreover, we will discuss suggestions for future research. Lastly, we make concluding remarks regarding the work presented in this thesis and the current research in renal cell carcinoma (RCC).

Summary of the main findings:

In this thesis, we assessed whether various environmental and genetic risk factors are differentially associated with the risk of specific subgroups of renal cell carcinoma. In **chapter 2**, we observed that the association between type 2 diabetes mellitus and Renal Cell Carcinoma (RCC) risk is present in women, but not in men. Upon further analysis, individuals who reported the use of anti-diabetic medication, in particular insulin and its analogues, had an increased risk of RCC (**chapter 2**). Furthermore, we found evidence for heterogeneity of associations regarding body mass index (BMI) and the risk of clear cell RCC (ccRCC) and papillary RCC (pRCC), as BMI was associated positively with ccRCC risk, but inversely with pRCC risk (**chapter 3**). No clear evidence was found for heterogeneity of associations regarding the association between cigarette smoking, hypertension and antihypertensives, and alcohol consumption and the risk of ccRCC and pRCC. In **chapter 4**, we investigated the association between a history of kidney stones and the development of RCC and upper tract urothelial carcinoma (UTUC). Having a history of kidney stones was associated with both an increased RCC and UTUC risk. Having a younger age, before 40 years of age, at first diagnosis of kidney stones increased these risks, when compared to a later diagnosis. In particular, there was a potential heterogeneity of associations as having a history of kidney stones was associated with an increased pRCC risk, but not with clear cell RCC risk. No heterogeneity of associations was observed between kidney stones and different UTUC localisations. In **chapter 5** we investigated the role of some selected germline single nucleotide polymorphisms (SNPs) in *VHL* and *HIF1A* on (cc)RCC risk. Overall, *VHL*_rs779805 was associated with the risk of RCC overall and ccRCC, while no associations were found for two other *VHL* SNPs and a *HIF1A* SNP. In addition, we assessed gene-environment and gene-gene interactions for these *VHL* and *HIF1A* SNPs. There was a potential gene-environment interaction between alcohol consumption and *VHL*_rs1642739, *VHL_rs779805* and *HIF1A_rs2301111* in RCC, but these interactions did not maintain statistical significance after correcting for the false discovery rate. No gene-gene interactions were found between the selected SNPs. Finally, we used targeted sequencing to examine the association between somatic mutations in the seven most frequently mutated genes in ccRCC, namely *VHL, PBRM1, SETD2, BAP1, MTOR, KDM5C and TP53,* and ccRCC-specific survival (**chapter 6**). Mutations in *VHL* and *PBRM1* were associated with ccRCC-specific survival, regardless of co-occurrence. However, these results did not maintain statistical significance after multiple testing correction.

Methodological considerations

The use of observational and molecular data, as included in this thesis, may introduce various sources of bias, which may have influenced the validity of our findings. This section will highlight and discuss various methodological strengths and limitations, and the influences these strengths and limitations may have on the results presented in this thesis.

General considerations
There were several important strengths regarding the analyses in this thesis, including the longitudinal study design, the availability of extensive confounder information, and the completeness of follow-up. Compared to case-control studies, the more commonly employed observational study designs in RCC research, prospective cohort studies have a reduced risk of information and selection bias. This, because the exposure information has been determined prior to the assessment of the outcome and all participants were at risk of developing the disease upon inclusion in our longitudinal study. Moreover, in the Netherlands Cohort Study on diet and cancer (NLCS) extensive information was available on potential confounding factors, which enabled us to extensively adjust for confounders in our analyses. Lastly, the completeness of follow-up for cancer incidence through record linkage is estimated to be over 96%[1].

Use of the self-administered baseline questionnaire
A methodological consideration that should be taken into account when interpreting the analyses presented in this thesis is the use of exposure data obtained through the use of a self-administered baseline questionnaire. These data were used in **chapter 2** through **chapter 5** to determine the exposure to various risk factors, including among others cigarette smoking, body mass index, hypertension, alcohol consumption, diabetes mellitus and kidney stones. A limitation is the absence of repeated measurements of exposure status during follow-up due to the design of the NLCS. Therefore, we need to consider two factors. Namely the potential for misclassification of exposures through self-report and the potential for change in exposure status in the period beyond baseline.

The misclassification of exposures may have occurred as the baseline questionnaire was based on self-report without additional verification of exposure. Therefore, the misclassification of exposure status may be a potential source of information bias. An advantage is the prospective nature of the Netherlands Cohort Study in which the participants were all at risk of developing cancer at the onset of the study. As such, it is unlikely that information bias has occurred, as a systematic difference in the accuracy of reported information between the randomly selected subcohort and participants who developed renal cell cancer is not expected. However, non-differential misclassification may have occurred. In general, non-differential misclassification most likely leads to the attenuation of results towards the null for binary exposures[2]. In the case of non-binary exposures, or upon collapsing continuous exposure data into categorical data, the effects of the misclassification of exposures is more difficult to predict[2].

Moreover, the measurement of exposure data at one point in time may be a source of potential bias, as we do not have information on the change in exposures beyond the baseline measurement. We believe that the participants in the age group as targeted in the NLCS, may have had relatively stable lifestyle habits over a period of at least 5 years after completing the baseline questionnaire, as reflected during reproducibility studies regarding the food frequency questionnaire (FFQ) included in the NLCS questionnaire[3, 4]. However, there is a possibility that participants may have changed their habits at some point or that they may have been diagnosed with medical conditions during follow-up. To gain more insight into this, we have assessed the likely direction of misclassification, based on evidence from other studies,

for the main risk factors derived from the baseline questionnaire in this thesis. Regarding (cigarette) smoking, we would expect that the proportion of never smokers would not be heavily affected, as very few people start smoking at older ages[5]. On the other hand, based on information from a Dutch study investigating the change in lifestyle habits, we expect that approximately 8%-12% of the individuals classified as current smokers at baseline may have quit smoking during follow-up[5]. This may have led to an underestimation of the effect of cigarette smoking on RCC subtypes in our study. Moreover, BMI may also have varied over time beyond baseline. The subcohort of the NLCS has reported their body weight (kg) at three separate occasions at baseline in 1986 and during follow-up in 1992, and 2000[6, 7]. In addition, the self-reported height was available from the baseline questionnaire in 1986. Using this information, the change of BMI until 14 years of follow-up could be determined. BMI did not increase over time in men, but increased slightly over time in women (0.2 kg/m^2 and 0.4 kg/m^2 from baseline to 6 and 14 years of follow-up, respectively)[7]. Based on these small increases in women, we expect that the associations for BMI should not have been affected by changes in BMI beyond baseline. Moreover, sharp increases in the prevalence of hypertension and hypertension with anti-hypertensive medication at older ages have been observed after repeated measurements in two cohorts[5, 8]. The prevalence of hypertension was found to increase with more than 20% in both men and women aged 50-59 years over a period of 11 years[5]. Increases were also observed for individuals with hypertension who used antihypertensive medication in Germany, these increases were the largest in age categories of 55-64 years (29%) and 65-74 years (22%) over a follow-up period of 9 years on average[8]. Based on these changes over time, we assume that similar trends may have occurred in the NLCS, which would likely lead to an attenuation of results in our analyses, in particular due to changes in hypertension status in the younger age groups. In the NLCS, questions were included in the baseline questionnaire to detect differences over time with regards to the consumption of alcohol before baseline. This information was used to assess whether individuals were stable alcohol users or abstainers during the 5-year period preceding baseline and separate analyses have been performed using this information. However, this does not provide information on the change in alcohol consumption during follow-up. As elderly tend to experience more severe and more prolonged effects of drinking alcohol, they may potentially reduce their consumption at older ages[9]. Such an effect was observed in a 20-year follow-up of a community sample from the United States, in which a reduction was observed in both the proportion of alcohol consumers and the frequency of consumption[10]. Another study, which showed similar trends, indicated that the consumption of lower quantities (one drink) increased with old age[11]. In light of these studies, the true (inverse) association between alcohol consumption and RCC risk may in reality be stronger than the association observed in our study, as individuals potentially decreased their alcohol consumption over time and more often consumed less alcohol than what they reported in the baseline questionnaire. We also used information on the self-reported history of kidney stones as diagnosed by a physician from the baseline questionnaire. In general, as the passing of a kidney stone generally is a memorable event, we expect that the reporting of kidney stones was accurate. In addition, the peak incidence of kidney stones is at 45-49 years of age, with the incidence decreasing after 55-59 years[12]. Noteworthy however, is the rise in kidney stone prevalence over the years, which may indicate that the increased occurrence of comorbidities implicated in kidney stone formation (among others, diabetes mellitus and hypertension), may have contributed to the more frequent occurrence of kidney stones during

follow-up. The diagnosis of kidney stones during follow-up likely indicates that the true effect of kidney stones on RCC development is larger than observed in our study. Regarding the ascertainment of diabetes mellitus in our cohort, we expect that more participants may have been diagnosed with type 2 diabetes mellitus during follow-up. In the years beyond baseline, there was a greater alertness regarding diabetes mellitus as a public health problem, largely due to the implementation of the *Standard Diabetes Mellitus Type II* in 1988, which contained guidelines on diagnosis, treatment and support of diabetes patients[13]. Furthermore, the Hoorn study which assessed the diabetes status of 2472 individuals aged 50-75 years in the Dutch town of Hoorn in the period 1989-1992 indicated that approximately 50% of all patients with diabetes were undiagnosed[14]. This, combined with the increased alertness for diabetes mellitus, increases the likelihood that additional diagnoses will have been made during follow-up by general practitioners. Furthermore, an increase in prevalence of approximately 75% was observed in the time-period 1991-2000 in men, and 25% in women in the Netherlands[15]. This change may largely be attributed to changes in the prevalence of risk factors for diabetes mellitus in the Netherlands (among others obesity, physical inactivity and smoking)[15]. In the period beyond 1999, changes have been made in the diagnostic criteria for diabetes mellitus, which may have contributed to a rise in prevalence beyond 1999[15]. Although the effects of this change will likely be small, as this affects only the last 7 years of follow-up. Overall, it is likely that the prevalence of diagnosed diabetes mellitus will also have increased during follow-up in our cohort, which will most likely have led to the underestimation of effects in the relationship between type 2 diabetes mellitus and RCC in our study.

Selection bias in the collection and sequencing of tumour blocks
Analyses using the NLCS data are generally unlikely to be influenced by selection bias, as participants who reported cancer at baseline were excluded from the analyses and the loss to follow-up was limited. However, selection bias may have occurred during the collection of tumour blocks of RCC cases, during the selection of samples with sufficiently available isolated DNA for performing targeted sequencing and the selection of samples as a result of the quality of sequencing as described in **chapter 6**[16].

The completeness of follow-up for cancer in the NLCS was high through the use of record linkage with the Netherlands Cancer Registry (NCR), the Dutch Pathology Registry (PALGA) and Statistics Netherlands (CBS). Two studies assessing the coverage of the NCR have indicated coverage rates of 96% in the Limburg region and in Rotterdam (upon restriction to pathology-confirmed cancers)[19, 20]. In the NLCS, at the onset of the study (1987) the coverage by the NCR alone was 89%[1]. Supplementing with information from PALGA helped attain a coverage of 98.5%, and in 1988 the proportion of cancer cases that could be retrieved through the combined efforts of the NCR and PALGA was 100%[1]. These findings highlight the completeness of cancer follow-up in our study and as a result the accuracy of the estimated person-time at risk was high. Therefore, we estimate that selection bias by loss to follow-up was minimal.

The potential for selection bias related to the collection of tumour blocks has been described in previous publications within the NLCS[17, 18]. In total, formalin-fixed paraffin-embedded tumour material of 454 cases out of 568 histologically confirmed epithelial RCC cases was

able to be collected (80%) from ~50 pathology laboratories throughout the Netherlands. Some reasons for the unavailability of tumour blocks were the absence of tumour material during the collection process (e.g. stage IV tumours are often not excised), the inability to send the tumour tissue by pathology laboratories and the incomplete linkage with the Dutch Pathology Registry (PALGA), in particular for cases before 1991[17, 18]. One method to assess whether selection bias has occurred due to sample loss is comparing whether exposures are related to the presence or absence of tumour blocks[16]. There were no differences on various exposure factors including, among others, age, sex, BMI, smoking, alcohol intake and hypertension between incident RCC cases with and without successful collection of tumour blocks[17]. In addition to this, due to the exceptionally high retrieval of tumour material (80%), we consider the potential of selection bias regarding the collection of RCC tumour blocks to be low[17].

From the available ccRCC tumour blocks (n=366) tumour DNA from 252 ccRCC cases was selected for targeted sequencing as described in **chapter 6**. This selection was based on the availability of sufficient DNA required for sequencing and the presence of DNA fragments of at least 200 base pairs based on a DNA ladder. This selection was important as severely fragmented DNA, which is a common occurrence in routinely archived formalin-fixed paraffin-embedded (FFPE) tissue, is less suitable for sequencing. After sequencing, 121 samples had an average unique read depth of at least 20x for six out of seven selected genes. To assess the potential of selection bias we assessed whether there were differences based on the distribution of clinical characteristics (*i.e.* age at diagnosis, sex, tumour grade, tumour stage and tumour size) between all ccRCC cases with available tumour blocks (n=266), cases eligible for targeted DNA sequencing (n=252) and samples included in the prognostic analyses in **chapter 6** (n=121). We assessed clinical characteristics, instead of environmental exposures as we did for the assessment of selection bias based on tumour collection, because these clinical characteristics may be associated with the prognosis of ccRCC, which was the outcome under study. In addition, these variables were available for all tumour blocks, based on the histological revision by two genitourinary pathologists. We did not observe clear differences between the distribution of clinical characteristics between these different selection steps. Therefore, we also consider the potential of selection bias to be low regarding the selection and inclusion of samples for targeted sequencing.

Histologic revision and the potential for misclassification of tumours
The classification of collected FFPE tumour material was performed by centralised revision of haematoxylin and eosin (HE)-stained slides by two experienced genitourinary pathologists based on the WHO-classification of tumours of 2004[21]. Based on this revision of tumour material of 454 RCC cases, 366 (80.6%) were clear-cell (cc)RCC cases, 60 (13.2%) papillary (p)RCC cases, 15 (3.3%) chromophobe RCC cases, and 13 (2.9%) other or undefined RCC cases. The collection and histological revision of tumour samples is a unique asset of the NLCS, which provides a means to investigate relationship between different histologic entities and the aetiology and prognosis of cancer. In addition, the centralised revision by two genitourinary pathologists allows for a more accurate representation of tumour histology, as large differences may be expected if the histologic revision was based on information as entered in the NCR or derived from separate pathologists at the ~50 laboratories of tumour block origin over the course of more than 20 years of follow-up.

However, due to recent advances and changes in the classification of RCC tumours there are potential differences in tumour classification that may be present if the tumour blocks would be revised using present day classification systems. These advances include the emergence and ascertainment of new renal entities in the time-period after the pathological revision in the NLCS was performed[22, 23]. These newly described emerging renal tumour entities are relatively rare, compared to ccRCC and pRCC, with the estimated occurrence per subtype ranging from <1% to 4% of all (adult) RCCs[24]. Of these new entities, clear cell papillary RCC (ccpRCC) is the most frequent entity, with an occurrence of approximately 3-4%[24, 25]. Noteworthy for ccpRCC is the overlap in morphological features with low-grade ccRCC and pRCC, which increases the difficulty to accurately distinguish this subtype[24, 26]. Therefore, histological confirmation of ccpRCC often necessitates the assessment of immunohistochemical characteristics for accurate classification[24]. As these subtypes were not yet included as major subtypes in the WHO classification in 2004, and the methods of accurately assessing these subtypes were yet to be standardised, these new subtypes were not included in our classification[21]. As a result, our histologic classification likely contains ccpRCC cases that were categorised as either ccRCC or pRCC. It is likely, as indicated in this thesis, that different renal entities possess different aetiologies. Therefore, the presence of ccpRCC cases in subgroups categorised as ccRCC or pRCC may have distorted our results. We estimate that the effect of this difference in classification would be most pronounced in pRCC cases as these were limited in number in our analyses and, in turn, the misclassification of cases could then have a larger effect. However, based on the rare occurrence of these new entities, when compared to the main histological subtypes, we estimate that the impact of this potential difference in classification will have been limited.

In addition to changes in the histological classification of tumours, changes have also been made in the UICC TNM classification of malignant tumours during the follow-up period of the NLCS. These changes have been integrated in the Netherlands Cancer Registry (NCR) during the follow-up. The most impactful changes were alterations made in the classification between T1 and T2 regarding the tumour size, with the original cut-off between 'small' and 'large' tumours, used by the NCR until 1988 (TNM: third edition[27]), <2.5 and ≥ 2.5 cm used by the NCR until 1998 (TNM: fourth edition[28] through the 2nd revision of the fourth edition[29]), and a cut-off threshold between T1 and T2 at 7 cm in the time-period beyond that (TNM: fifth edition and up[30, 31]). In addition, different versions were not comparable regarding the lymph node involvement and changes have been made regarding the classification of the number and dimension of involved lymph nodes, with increasing strictness across later versions[27-31]. To remedy potential differences in classification over time, an older tumour grade classification (Fuhrman grading system) and all tumours in the NLCS were recoded according to the third edition of the TNM classification, as used in **chapter 6**[27]. To remedy the differences in T-stage, we adjusted all our models that included tumour stage with tumour size.

Considerations in molecular (epidemiologic) research

Molecular research is a rapidly evolving field in which new revolutionary technologies are introduced in quick successions to uncover new genetic markers for disease[32]. Indeed, with the use of sophisticated next generation sequencing technologies and the continuous reduction of sequencing costs ample opportunities exist to unravel the biological complexities of the human genome. However, based on the aim of a study, various analysis methods exist

ranging from the identification of variants in candidate-genes to the use of next-generation sequencing techniques (NGS).

The detection of potential risk loci in RCC for the study of gene-environment interactions
In present day cancer research, the identification of novel risk loci is largely driven by the use of genome-wide association studies (GWAS). In these GWAS studies hundreds of thousands of Single Nucleotide Polymorphisms are tested for associations in an agnostic manner after which stringent genome-wide significance p-value thresholds are applied. As a result, only the most consistently and (relatively) strongly associated SNPs are maintained with odds ratios ranging from 1.1-1.4 in RCC[33-40]. At present, at least 13 autosomal risk loci have been described in the development of RCC and at least 2 sex-specific loci have been reported[33-40]. The GWAS approach stands in strong contrast to the use of a more traditional candidate gene approach, in which often specific pathways of interest are analysed in a hypothesis-driven manner, with specific associations between genes and the environment in mind[41]. With the advent of GWAS studies, due to the increased availability of SNP-chips and the limited power and low replicability of the often smaller candidate gene studies, the candidate gene study has taken a backseat in genetic research with some researchers even hinting at the obsoleteness of candidate gene studies in genetic research[42]. However, these GWAS studies do not come without risks either, as there are some methodological considerations that need to be made. Firstly, genetic traits identified in large GWAS studies are often determined based on linkage disequilibrium in the genome, which is the non-random linkage of genetic variants that are inherited together, indicating that a variant or a linked non-measured variant is associated in an independent model on disease outcome. As a result, it is often difficult to draw clear inferences from the findings in GWAS to make the step from sequence to consequence and from risk locus to involved gene[43]. In that sense, there is a niche opportunity for candidate gene studies to assess risk polymorphisms at a deeper level, when compared to GWAS studies. In candidate gene studies, as performed in **chapter 5**, it is important to critically assess potential polymorphisms based on *a priori* hypotheses. In the research described in this thesis, three SNPs in *VHL* were selected based on their association with either RCC risk or *VHL* promoter methylation in prior research[44, 45]. In addition, one *HIF1A* tag-SNP was included to investigate the potential for gene-environment interactions between SNPs in *HIF1A* and RCC, as the VHL/HIF-pathway is known to be involved in the aetiology of RCC[46]. Indeed, in **chapter 5**, an association was found between one of the included *VHL* SNPs and (cc)RCC risk in our models. While this SNP is readily available on commonly employed SNP assays used in GWAS studies, it has not yet been detected as a potential risk locus in GWAS studies, even though it has been consistently detected in multiple candidate gene studies[44, 47-49]. This indicates that candidate gene studies can still be relevant for the detection of plausible risk loci, and that risk loci should not solely stem from findings from large GWAS studies[50]. However, before performing a candidate gene study, serious considerations need to be made. First and foremost, sample sizes and biological plausibility should be critically assessed to be able to detect the presence of the relative small increases in risk conveyed by SNPs in relevant areas of the genome. Furthermore, information on the biological plausibility can be derived from either mechanistic hypotheses due to the involvement of a key pathway in tumourigenesis, or it could stem from novel regions of interest as detected in GWAS studies to try and pinpoint risk SNPs in potential functional elements (e.g. regulatory sequences such as promoters)[51]. Moreover, one of the biggest

criticisms of candidate gene studies is the non-replication of results, which is often thought to be attributed to population stratification[52]. Therefore, candidate gene studies should focus on extensive validation of detected associations in separate studies and populations to increase the reproducibility. In addition, even though often few candidate SNPs are tested, candidate gene studies require the implementation of false discovery rate procedures as there is a considerable risk of false discoveries. While genome-wide significance p-values are often too stringent, other techniques such as the (adaptive) Benjamini-Hochberg method provide a valid method for controlling false discoveries in candidate gene studies[53]. Upon taking these factors into account a niche might be created for candidate gene studies to provide valuable insights alongside the evidence generated by genome-wide association studies.

The selection and use of (NGS) techniques in this thesis
There are several NGS-based techniques available for the use in molecular research. In **chapter 6** of this thesis we describe a study in which a target panel of the coding regions of 42 genes was sequenced using a targeted sequencing approach. This approach was chosen to optimize the number of samples available for analysis, taking into account the costs when compared to whole genome sequencing (WGS) and whole exome sequencing (WES). In addition, the currently employed targeted sequencing panel enabled us to focus on setting variant call thresholds, to reduce the time needed to interpret the data, and it allowed for the visual verification of mutations in regions of interest. More comprehensive techniques, such as WES and WGS, result in long lists of variants with (yet) unknown significance[54]. As a result, the clinical relevance of these results will be difficult to assess. Therefore, we believe the chosen method was the most feasible, taking into account the means available at the onset of this study.

At the onset of this study, we focused on the sequencing of 32 of the most frequently mutated genes in ccRCC, supplemented with 10 genes associated with the VHL/HIF pathway and the P13K/AKT/MTOR pathway in ccRCC[55-57]. The original aim of this research was to assess somatic mutations, pathways and clustered analyses. However, because of the limited number of samples with sufficient sequencing quality (n=121), related power issues when assessing genes with a lower mutation prevalence (<5%), we restricted the analyses as presented in **chapter 6** to the seven most frequently mutated genes in ccRCC (i.e. *VHL, PBRM1, SETD2, BAP1, MTOR and TP53*)[55, 56].

Suggestions and potential pitfalls for the use of sequencing data in large cohort studies
The collection and storage of fresh-frozen tissue, which is considered to be the standard for use in NGS, is often not feasible in large cohort studies due to the difficulty in logistics and costs of storage. Therefore, routinely archived FFPE tissue, as was collected in the NLCS, might pose an alternative to fresh-frozen samples, when taking into account the easy handling, long-term cheap storage and accessibility. While the fixation delay, the fixation process, tissue preparation, paraffin embedding and archival storage are known to contribute to DNA fragmentation, cross-linking and chemical modification of FFPE tissue-derived DNA, FFPE tissue derived DNA can provide valuable information regarding mutations in tumour DNA[58]. Unfortunately, the aforementioned factors will likely have biased the DNA quality of samples in this thesis, as we were not able to discern between pathology laboratories and the associated fixation practices and storage conditions per laboratory in the

analyses. Therefore, selective drop out of samples may have occurred during the sequencing process based on differences in practices and guidelines between pathology laboratories. In addition, fixation and storage practices in Dutch pathology laboratories may have changed over time and small variations in fixation protocols may have led to differences in quality and quantity of isolated DNA[59].

While we were unable to assess the fixation practices in the NLCS, we did use the available information on the date of diagnosis and the date of DNA isolation to obtain information on the storage duration of FFPE blocks. In addition, we had information on the DNA stock concentration after DNA isolation, DNA fragment length as determined by a DNA ladder, and the estimated percentage of tumour tissue per tumour block. We used this information to assess which of these factors may have contributed to the quality of sequencing, operationalised as average read depth, of the 252 samples sequenced in the NLCS. Overall, we observed that samples with a longer storage duration as a FFPE block, a lower DNA stock concentration and higher fragmentation prior to sequencing had a lower average read depth. No direct association was observed between the percentage of tumour content and average read depth. Consequently, we noted that samples that had sufficient DNA quality to be included in analyses in this thesis had a shorter storage duration, a higher stock DNA concentration and a longer fragment length prior to sequencing. Two previous studies have found similar evidence regarding the association between storage duration and the reduction of sequencing quality[58], while other studies found that storage up to 12 and 18 years only had a minor impact on the sequencing quality[60, 61]. Secondly, the long-term storage of DNA may negatively impact the stability and degradation of DNA when stored in low concentration[62]. This could explain the reduced sequencing quality observed in samples with reduced DNA stock concentrations during storage. However, it is important to note that the input quantity of DNA for sequencing was standardised for all samples under study. Lastly, DNA fragmentation is a substantial and well-documented problem in FFPE samples[63, 64]. Therefore, assessing which samples have limited DNA fragmentation prior to sequencing is valuable when selecting which samples to sequence as NGS fails more often when DNA is extracted from old FFPE samples[58].

While these findings by themselves do not completely pinpoint the reason why a large proportion of samples have dropped out in our study, they still may hold great value for future large scale cohort studies willing to obtain DNA from samples to manage expectations and to assess potential measures to counteract the temporal changes in the sample DNA. In general, FFPE tissue has been indicated as a source of robust data with comparable results to fresh-frozen tissue[61, 65, 66]. Findings reported in this thesis indicate that extensive quality control prior to sequencing may provide valuable opportunities for enhancing NGS results when using routinely archived FFPE tissue and hopefully may lead to lower drop-out rates of sequenced samples.

The selection of variant calling thresholds

In the sequencing process, various decisions have been made to optimise the variant calling process of the NGS data. Firstly, two variant callers, namely Freebayes and HaplotypeCaller from the Genome Analysis Toolkit (GATK), were employed to detect single nucleotide variants, insertions and deletions in our study. One limiting factor in this study was the absence of matching normal tissue to the tumour tissue, which increased the difficulty in the identification of true somatic mutations. To increase the likelihood of detecting somatic mutations, variants with a population-based frequency over 1% were filtered out. As it is likely that common population-specific variants were still present, we excluded variants that were present in more than four samples. FFPE tissue is also known to contain random C>T and G>A sequence artifacts[67]. By using a molecular barcoding-based method, which enabled us to detect unique sequencing reads, we likely reduced the likelihood that we detected these random artifacts as a major clone in our analyses[67]. In addition, we have assessed various methods of denoting which variants were major clones in our dataset and should be maintained for analysis, including changing the threshold regarding the mutant read frequency, minimum alternate reads and the required number of reads needed to call mutations. Major clones were defined as variants that were likely to be present in the majority of tumour cells. Due to the varying read depth between samples in our studies setting robust thresholds often led to the over- or insensitivity of the detection of variants in our study. As a result, gene mutation percentages were highly variable depending on the chosen thresholds. Therefore, a per sample proportional mutant read frequency approach was employed, in which the somatic variant with the highest mutation read frequency within the 42 included genes was called as a major clone. Subsequently, variants with a mutant read frequency ≥50% of the highest mutant read frequency seen for that sample and with at least four alternate reads were also detected as major clone variants. This also provided a method to deal with samples with a heterogeneous tumour content, which normally require adjusted variant detection limits depending on the tumour content percentage[68]. There are some limitations to the use of this method. Firstly, this method assumes that the coverage for all included regions is sufficiently high to reliably call mutations. Therefore, samples and target regions that did not function properly were assessed and excluded prior to the application of this method. Secondly, the assumption is made that at least one mutation is present in the 42 included genes under study, after earlier exclusion steps, which is then assessed as a major clone for that sample, unless no mutation is found with an alternative read frequency of at least four. Overall, the overall gene mutation frequency for the 121 samples were slightly lower compared to the TCGA and COSMIC databases as depicted in Table 1, except for *KDM5C*, which was slightly higher. Likely we were conservative in the calling of mutations, which is also exemplified in the lower variant calling rate with the currently employed method compared to prior studies in the NLCS[18, 69]. This may in part be attributable to the focus on major clones in this study, while other studies often included all detectable variants. Considering these factors, it is likely that the mutation percentages are lower than if we had decided to be less conservative with our variant calling. However, that would have likely led to an increase in false positive mutation calls.

Table 1 - Comparison in gene mutation frequencies between the NLCS, the TCGA and COSMIC

	NLCS	TCGA (PanCancer)[a,70]	COSMIC[b, 56]
VHL	37%	41%	52%
PBRM1	22%	38%	30%
SETD2	12%	12%	13%
KDM5C	12%	5%	7%
BAP1	6%	10%	13%
MTOR	3%	8%	7%
TP53	3%	3%	7%

[a] Kidney Renal Clear Cell Carcinoma (TCGA), PanCancer Atlas – samples with mutation data – accessed 02-July-2020 through cbioportal.org[71, 72]
[b] accessed 02 July-2020 through https://cancer.sanger.ac.uk/cosmic

The potential impact of intratumour heterogeneity

At the time of the collection of tumour blocks, one tumour block was collected per case for the isolation of DNA. Therefore, the results derived from our sequencing efforts reflect the occurrence of mutations in one tumour block. However, in present day research there is more attention for the potential intratumour heterogeneity in cancer, in which different mutations are found in different segments of the tumour, or in our case, within one tumour block. In general, tumours evolve into multiple genetically distinct subclones, which follow a branched evolution reflecting the tumour's life history[73]. Mutations that are present in all tumour cells generally reflect mutations that have been acquired in the beginning of tumourigenesis. These major somatic mutations are thought to either be drivers or initiators of the process of tumourigenesis, or they may be passenger mutations that were already present in the cell prior to the transformation into a malignant cell[73]. In studies in ccRCC, inactivation of the *VHL* tumour suppressor gene and loss of heterozygosity at chromosome 3p were seen as truncal early events[74, 75]. In addition, *PBRM1* was also commonly observed as a truncal driver mutation in ccRCC[74]. Other common mutations in expected driver genes in ccRCC, such as *BAP1, PTEN, PIK3CA, SETD2* and *TP53*, are often subclonal and play a role in progression of the tumour[74]. Of interest is the parallel evolution occurring in different subclones of *SETD2, BAP1, KDM5C, ARID1A* and *PBRM1*. These genes are recognised as chromatin modifiers, regulators of genomic architecture and DNA accessibility, which are important for gene expression and DNA damage repair, and have been consistently detected in subgroups of cases in single-sample analyses of ccRCC[75, 76]. Aside from the mechanical and clinical implications of these genes in tumour progression, this may also be an important consideration for the results of this thesis. As we only assessed one tumour block from our patients, we may only have a limited insight in the overall mutational profile of these patients. While it is likely that we have observed the most important driver mutations, as we commonly observed mutations in *VHL* and *PBRM1*, we were only able to detect mutations that were present in the specific area of origin of the tumour block. Resultingly, it is likely that we may not have been able to detect the full spectrum of mutations present in the (complete) tumour in our sequencing results. A solution to this, would be to introduce multiregional sampling, or alternatively, to collect multiple tumour blocks[75, 77]. Unfortunately, this is generally not feasible in large nationwide cohort studies due to the complicated logistics involved in the

selection tumour blocks or regarding the logistics and planning for the sampling of multiple areas in the tumour. In hindsight, an alternative could have been to collect and histologically revise multiple tumour blocks per individual. In turn, the tumour block with the highest tumour grade should then have been selected for sequencing, as tumour grade has been associated with mutational load in studies on intratumour heterogeneity[75, 78]. In addition, this allows for the identification of mutations in the most aggressive tumour cells. Potentially, this could have provided us with a more complete overview of the mutational burden for ccRCC in the tumours of the included samples. Future studies should ideally include protocols for multiregional sampling to provide crucial insights in the aetiology and prognosis of ccRCC, to assess tumour progression and to obtain a more complete snapshot of the mutational profile of the tumour.

Implications of the findings of this thesis

One of the aims of this thesis was to identify whether various environmental risk factors are differentially associated with the risk of renal cell carcinoma and its subtypes. With regards to environmental risk factors, we observed a potential heterogeneity of associations for BMI across ccRCC and pRCC risk. In the study detailed in **chapter 3**, BMI was positively associated with the risk of ccRCC, but inversely with pRCC risk. No clear heterogeneity of associations was found for cigarette smoking, alcohol consumption, hypertension and antihypertensive medication. The consistent reports of heterogeneity for BMI across ccRCC and pRCC risk may provide an indication of etiologic differences across RCC subtypes[79-81]. Until recently, from the other investigated risk factors, solely specific subtypes of antihypertensive medication have been heterogeneously associated across histological subtypes in other studies[82]. While we did assess the use of antihypertensive medication in our study, for which we observed no heterogeneity, we were unable to look into detail at the different types of antihypertensive medication due to power considerations and due to the use of different types of antihypertensive medication at the time. As a result, we were not able to validate these findings regarding specific types of antihypertensive medication. In a recently published study, which investigated the prevalence of cigarette smoking among patients with different histologic RCC subtypes, a higher prevalence of smoking was observed in pRCC, compared to ccRCC, although the difference was relatively small[83]. The extremely high sample size, due to the use of population-based cancer registries, may in part explain why this study was able to find differences in smoking status, while other smaller studies did not observe such heterogeneity.

These studies may act as a starting point to unravel differences in the mechanisms that influence the risk of specific RCC subtypes. There is a constant further sophistication in the classification of subtypes in RCC. Recent research has highlighted that there is a potential molecular subtyping present in ccRCC, designated as ccA and ccB, as originally described by Brannon *et al*.[84]. Of particular interest is the new finding that ccA risk may even be more strongly associated with BMI, when compared to ccB[85]. Moreover, the subclassification of pRCC into type 1 and type 2 in aetiology research may also provide worthwhile avenues for the detection of more specific aetiologic relationships in subtypes of RCC in future research. Papillary RCC is of particular interest, as type 1 and type 2 pRCC, also possess a distinct genomic profile. Type 1 pRCC more often features *MET* alterations, either through mutations or gain of chromosome 7 where *MET* is located[86]. Type 2 pRCC is characterised by a more

heterogeneous genomic profile, as it tends to have alterations in *FH, CDKN2A, TFE, TFEB, SETD2, BAP1 and PBRM1* genes[86]. Based on these differences, these pRCC subtypes could possess different mechanisms for developing cancer, and may as a result also be differentially associated with aetiologic risk factors. Unfortunately, in our studies, we were unable to discriminate between pRCC type 1 and type 2, even though the classification information was readily available, due to the limited number of pRCC cases in our study. These novel findings, combined with the evidence described in this thesis, highlight that the aetiology of ccRCC and pRCC is even more complex than previously understood.

While there is great focus on the established risk factors for RCC, other potential risk factors for RCC are still relatively poorly understood. In particular, evidence from large prospective cohort studies regarding these risk factors is lacking or needs confirmation. Therefore, we studied two potential risk factors of interest in this dissertation, namely type 2 diabetes mellitus (**chapter 2**) and the history of kidney stones (**chapter 4**).

In **chapter 2**, we assessed the association between type 2 diabetes mellitus and RCC. We were able to confirm associations, as described in the Nurses' Health Study and the Health Professionals Follow-up Study[87], as the association between type 2 diabetes mellitus and RCC was present in women, but not in men. We also assessed the relationship between RCC and antidiabetic medication and observed that individuals with type 2 diabetes mellitus who used insulin or analogues had a strongly increased risk compared to individuals without type 2 diabetes mellitus. However, we need to be reserved and cautious when interpreting these results. As a result of the increasing stratification of subgroups and the further specification of exposures the number of cases in subgroups has become increasingly low, and effect estimates have become increasingly imprecise. Therefore, we hope that future studies are able to further discern this relationship as the relationship between anti-diabetic medication, in particularly metformin, and cancer has been a source of debate in cancer research. Based on the evidence from the Nurses' Health Study and the Health Professionals Follow-up study, which found similar relationships regarding the association between type 2 diabetes mellitus and renal cell cancer, we might need to revisit the relationship between diabetes mellitus and its medication and the risk of RCC and its subtypes with the use of modern methods to ascertain the status of diabetes mellitus[87]. In addition, there is a sense of urgency with regards to untangling this relationship, as the worldwide prevalence of diabetes is projected to keep increasing in the coming 25 years[88].

In **chapter 4**, we describe the association between the history of kidney stones and the occurrence of RCC and UTUC. At the time of publication, solely case-control and retrospective cohort studies had described this relationship, as detailed in the extensive meta-analysis by Cheungpasitporn *et al.*[89]. In addition, there were few new studies investigating this relationship as eight out of nine studies included in the meta-analysis were published before the year 2000, and only three retrospective cohort studies were published after the publication of the meta-analysis on this relationship[89-91]. The analysis in the NLCS was able to describe new evidence on the relationship between kidney stones and RCC and UTUC risk. Most interestingly, we observed a heterogeneity of associations regarding the history of kidney stones and ccRCC and pRCC risk that had not been described before, namely kidney stones were associated with an increased pRCC risk, but not ccRCC risk. This unexpected

finding needs further confirmation, as both these histologic subtypes are thought to stem from the proximal convoluted tubule[92]. At this location it is thought to be unlikely for kidney stones to form[93, 94]. Therefore, the difference in these subtypes might imply that not the kidney stone formation, but another factor correlated with kidney stones may be associated with the difference in aetiology of RCC subtypes. In this research, we did not have information on the type(s) of kidney stones and the recurrence of kidney stones. Potentially, the association between different kidney stone types and the risk of ccRCC and pRCC may provide additional insights on the influence of specific components on the development of these subtypes. Hopefully, this may help elucidate the mechanisms behind this peculiar finding. In addition, information on the recurrences of kidney stones could provide interesting evidence as the results detailed an increased RCC risk at earlier ages of first kidney stone. Potentially, the earlier age at first kidney stone could be an indicator of having a longer period with recurrent stones. Aside from the effects of stone formation, tumour development could also be related to the presence of stone-forming salts and the interaction with fluids. Previous research in the NLCS has indicated that the interaction between sodium and fluid was associated with RCC aetiology, which are both important factors for the development of kidney stones[95]. Therefore, while the reason behind the heterogeneous association between kidney stones and RCC subtypes is unclear, it may provide a starting point for future research.

In **chapter 5**, we confirmed the association between one *VHL* SNP and (cc)RCC risk. *VHL_rs779805*, while not (yet) discovered in large GWAS studies in RCC, has been associated with RCC risk in multiple candidate-gene studies[47, 48]. In addition, we observed that this risk was increased in ccRCC, when compared to overall RCC. While there is a clear link between the presence of mutations and epigenetic alterations in *VHL* for the risk of ccRCC, these findings also indicate that a smaller and more frequently occurring polymorphism in *VHL* may increase the susceptibility to ccRCC. In our study, we were not able to discern any clear gene-environment and gene-gene interactions. Moreover, based on earlier publications on *VHL* polymorphisms, we had expectations that there may be a possible link between tumour-specific *VHL* promoter methylation and the occurrence of *VHL* SNPs[44]. However, we did not observe an association between *VHL* SNPs and *VHL* promoter methylation in our study. Furthermore, various gene-environment interactions have been observed in (cc) RCC[96-100]. We hypothesised that there might be potential for interaction between the most common risk factors for RCC and SNPs in the VHL/HIF-pathway. However, we observed no clear indication that there is a role of environmental factors in increasing or decreasing the susceptibility for ccRCC in the presence of *VHL* or *HIF1A* SNPs.

To assess high-risk alterations to the DNA, we used NGS to sequence a panel of frequently mutated genes in ccRCC (**chapter 6**). Determining the prognostic relevance of the seven most frequently mutated genes in ccRCC is a daunting task. There is a need for detailed clinical information, long-term follow-up data and the availability of either tumour material or isolated tumour DNA. In this dissertation, we describe the favourable association in individuals with *VHL* and *PBRM1* mutations with ccRCC-specific survival, when compared to individuals without mutations in *VHL* and *PBRM1*. Evidence on the association between *VHL* mutations and ccRCC-specific survival is inconsistent, as studies have reported both a better[101] and a worse ccRCC-specific survival associated with *VHL* mutations[102]. In addition, in an earlier study in the NLCS no association was observed[69]. These varying observations

indicate the difficulty of discerning the influence of these mutations on ccRCC prognosis. The association of overall survival with *PBRM1* mutations has been described in a large retrospective study within Memorial Sloan Kettering Cancer Center[103], but not clearly with cause-specific survival using earlier Memorial Sloan Kettering Cancer Center and TCGA data[101, 104, 105]. These findings indicate the need for new large-scale studies to create additional insight into these relationships, as the current evidence is not sufficient to be able to provide unambiguous recommendations to the clinic regarding the prognosis of tumours based on their molecular make-up.

At present, the bulk of the evidence regarding the association between somatic mutations and the prognosis of ccRCC is often obtained using large databases, such as the TCGA and COSMIC. While many reasons support the use of such elaborate programmes, there may be some intrinsic factors that may affect the generalizability of results. Most importantly, when assessing the TCGA-KIRC classification (n=537) in the NIH Genomic Data Commons the source of samples is dominated by three large centres, namely the Memorial Sloan Kettering Cancer Center (26%), the University of Pittsburgh (20%), and the MD Anderson Cancer Center (13%)[55, 106]. As younger patients are more likely to be treated and included in research in these centres, this may affect the clinical characteristics and the treatments received of the population in this database[107]. Therefore, results and mutation rates as found in these initiatives are potentially driven by cases coming from these large specialised U.S. cancer and academic institutes, potentially affecting the genetic profile observed in ccRCC. In addition, these patients may also have received different treatments, when compared to other institutions, as these institutes are counted among the best hospitals for cancer in the U.S. and are more prone to use cutting-edge therapies. Potentially, this may affect outcomes when assessing the relationship between mutations and prognosis. Therefore, assessing the prognostic value of somatic mutations in an independent cohort may provide valuable information. In addition, novel research on samples from new large-scale studies might aid in getting a clearer view of the genomic characterisation of ccRCC, as there are some deviations, as indicated by table 1, in mutation rates between large databases for some of the most frequently mutated genes in ccRCC. This is exemplified in currently published studies in which a large variability in the proportion of ccRCC cases with *VHL* mutations is detected (17-95%)[108].

Recommendations for future research

The results in this thesis indicate that there is potential for heterogeneity in the aetiology between RCC subtypes, in particular for BMI and kidney stones. While the current evidence has been consistent across studies, more studies are needed to further solidify the evidence. In addition, more research is needed to be able to adequately assess potential aetiologic factors taking into account the different RCC subtypes. Potentially, the effect of these risk factors may have been distorted by the inclusion of different RCC subtypes in analyses on overall RCC in past research. While the number of cases in individual studies for pRCC, chromophobe RCC and other more infrequent subtypes are in general very limited, large collaborations or the presentation of results for use in large meta-analyses may help circumvent this methodological problem. At present, little is known about the exact mechanisms how, and why these aetiologic factors lead to differences in the risk of histologic subtypes of RCC. In the past, attempts have been made to tie various risk factors together with the most common mutations in RCC using hypotheses on the physiologic role of the kidney in the detection of

hypoxia[109]. However, a clear causal link between risk factors, hypoxia and RCC has not yet been described, aside from the involvement of the *VHL/HIF* pathway.

To create more insight into these mechanisms, future RCC research should focus on combining information from lifestyle and medical conditions, somatic mutations and epigenetic alterations. Assessing the co-occurrence of these factors requires methods that are able to agnostically assess combinations between environmental and genetic risk factors. At present, so-called mutographs are employed to assess the interplay between risk factors and alterations in the DNA, as described in the paper by Alexandrov *et al.*[110]. Using this method, unique mutational signatures are created to detect potential exposures to mutagens. With the help of specific mutational patterns, it may be possible to pinpoint the effect a specific exposure or activity exerts on the genome[111]. At present, more than 70 signatures have been identified in cancer, which have been associated with various proposed aetiologies including processes related to, among others, tobacco smoking and alcohol consumption[110]. In addition, signatures have been found that were specific to kidney papillary cancers[110]. However, the majority of mutational signatures remain of unknown cause[110]. In future research, underlying processes or tumour subtypes may be uncovered by assessing tumour profiles using more accurate exposure information or more elaborate mathematical approaches to derive the origin of mutations. Furthermore, future studies may also be able to use clustering techniques to determine common co-occurrences of mutations, copy number variations and variations in expression, as described in Begg *et al.*[112]. Methods like these may be able to detect combinations of aetiologic factors for cancer or specific cancer subtypes that would not readily be detected using more traditional methods. Unfortunately, the use of these mutographs and clustering techniques was beyond the scope of this thesis due to the need of more extensive genomic information than what is currently available within the Netherlands Cohort Study on diet and cancer (NLCS).

Concluding remarks

Renal cell carcinoma is complex due to the great amount of heterogeneity present between histologic subtypes. Continuous efforts are made to detail the underlying pathology with regards to risk factors, subtype classification and genetic alterations. Only by striving to classify renal cell carcinoma as accurately as possible can we get more insight on the mechanisms that drive this cancer. This will require epidemiologists to be highly adaptable to the continuously changing classifications or to critically assess the magnitude of these alterations on the results if older classifications are used.

The subclassification of renal cell carcinoma is not without problems. With continuously changing classifications and new insights on the genetic and histologic characteristics of RCC it is difficult to stay up-to-date with the present status of RCC research in large prospective cohort studies. Which often necessitates falling back on older classification systems to unify the available data collected over long follow-up durations. In addition, the constant stratification of subtypes in RCC leads to a decrease of power in analyses, which leads to difficulties in obtaining consistent results. Therefore, transparency in results and the combined effort of multiple research teams is required to enable future researchers to place findings in the right perspective.

With this dissertation, I hope to contribute by providing insight and critical considerations in the field of renal cell carcinoma research. In short, we observed heterogeneity of associations between ccRCC and pRCC regarding BMI and kidney stones. In addition, we confirmed findings regarding the association between kidney stones and type 2 diabetes mellitus with RCC. Furthermore, we observed differences in ccRCC risk between individuals with a *VHL* SNP. Lastly, using targeted sequencing, we observed that cases with mutations in *VHL* and *PBRM1* had a better prognosis, when compared to cases without mutations in these genes. However, these findings did not remain significant after correction for multiple testing.

References:

1. Goldbohm RA, Van den Brandt PA, Dorant E. Estimation of the coverage of Dutch municipalities by cancer registries and PALGA based on hospital discharge data. Tijdschr Soc Gezondheidsz. 1994;72: 80-84.
2. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Lippincott Williams & Wilkins, 2008.
3. Goldbohm RA, van 't Veer P, van den Brandt PA, et al. Reproducibility of a food frequency questionnaire and stability of dietary habits determined from five annually repeated measurements. Eur J Clin Nutr. 1995;49: 420-429.
4. Goldbohm RA, van den Brandt PA, Brants HA, et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. Eur J Clin Nutr. 1994;48: 253-265.
5. Blokstra A, Smit HA, Verschuren WMM. Veranderingen in leefstijl- en risicofactoren voor chronische ziekten met het ouder worden: De Doetinchem Studie 1987-2002 Bilthoven: RIVM, 2006.
6. Hughes LA, Arts IC, Ambergen T, et al. Higher dietary flavone, flavonol, and catechin intakes are associated with less of an increase in BMI over time in women: a longitudinal analysis from the Netherlands Cohort Study. Am J Clin Nutr. 2008;88: 1341-1352.
7. Gilsing AM, Weijenberg MP, Hughes LA, et al. Longitudinal changes in BMI in older adults are associated with meat consumption differentially, by type of meat consumed. J Nutr. 2012;142: 340-349.
8. Lacruz ME, Kluttig A, Hartwig S, et al. Prevalence and Incidence of Hypertension in the General Adult Population: Results of the CARLA-Cohort Study. Medicine. 2015;94: e952-e952.
9. Barry KL, Blow FC. Drinking Over the Lifespan: Focus on Older Adults. Alcohol Res. 2016;38: 115-120.
10. Moos RH, Schutte KK, Brennan PL, Moos BS. Older adults' alcohol consumption and late-life drinking problems: a 20-year perspective. Addiction. 2009;104: 1293-1302.
11. Breslow RA, Smothers B. Drinking patterns of older Americans: National Health Interview Surveys, 1997-2001. J Stud Alcohol. 2004;65: 232-240.
12. Romero V, Akpinar H, Assimos DG. Kidney stones: a global picture of prevalence, incidence, and associated risk factors. Rev Urol. 2010;12: e86-96.
13. Ruwaard D, Gijsen R, Bartelds AI, Hirasing RA, Verkleij H, Kromhout D. Is the incidence of diabetes increasing in all age-groups in The Netherlands? Results of the second study in the Dutch Sentinel Practice Network. Diabetes Care. 1996;19: 214-218.
14. Mooy JM, Grootenhuis PA, de Vries H, et al. Prevalence and determinants of glucose intolerance in a Dutch caucasian population. The Hoorn Study. Diabetes Care. 1995;18: 1270-1273.
15. Volksgezondheidszorg.info. Trend jaarprevalentie en nieuwe gevallen diabetes mellitus 1991-2014: https://www.volksgezondheidenzorg.info/onderwerp/diabetes-mellitus/cijfers-context/trends, RIVM: Bilthoven, 23 July 2020
16. Hoppin JA, Tolbert PE, Taylor JA, Schroeder JC, Holly EA. Potential for selection bias with tumor tissue retrieval in molecular epidemiology studies. Ann Epidemiol. 2002;12: 1-6.
17. Deckers IA, van Engeland M, van den Brandt PA, et al. Promoter CpG island methylation in ion transport mechanisms and associated dietary intakes jointly influence the risk of clear-cell renal cell cancer. Int J Epidemiol. 2017;46: 622-631.
18. van Houwelingen KP, van Dijk BA, Hulsbergen-van de Kaa CA, et al. Prevalence of von Hippel-Lindau gene mutations in sporadic renal cell carcinoma: results from The Netherlands cohort study. BMC Cancer. 2005;5: 57.

19. Schouten LJ, Höppener P, van den Brandt PA, Knottnerus JA, Jager JJ. Completeness of cancer registration in Limburg, The Netherlands. Int J Epidemiol. 1993;22: 369-376.
20. van der Willik KD, Ruiter R, van Rooij FJA, et al. Ascertainment of cancer in longitudinal research: The concordance between the Rotterdam Study and the Netherlands Cancer Registry. Int J Cancer. 2020;147: 633-640.
21. Eble J, Sauter G, Epstein J, Sesterhenn I. World Health Organization Classification of Tumours. Pathology and Genetics. Tumours of the Urinary System and Male Genital Organs. Lyon: IARC Press, 2004.
22. Mikami S, Oya M, Mizuno R, et al. Recent advances in renal cell carcinoma from a pathological point of view. Pathol Int. 2016;66: 481-490.
23. Udager AM, Mehra R. Morphologic, Molecular, and Taxonomic Evolution of Renal Cell Carcinoma: A Conceptual Perspective With Emphasis on Updates to the 2016 World Health Organization Classification. Archives of Pathology & Laboratory Medicine. 2016;140: 1026-1037.
24. Inamura K. Renal Cell Tumors: Understanding Their Molecular Pathological Epidemiology and the 2016 WHO Classification. Int J Mol Sci. 2017;18.
25. Chen WJ, Pan CC, Shen SH, et al. Clear cell papillary renal cell carcinoma - An indolent subtype of renal tumor. J Chin Med Assoc. 2018;81: 878-883.
26. Dhakal HP, McKenney JK, Khor LY, Reynolds JP, Magi-Galluzzi C, Przybycin CG. Renal Neoplasms With Overlapping Features of Clear Cell Renal Cell Carcinoma and Clear Cell Papillary Renal Cell Carcinoma: A Clinicopathologic Study of 37 Cases From a Single Institution. Am J Surg Pathol. 2016;40: 141-154.
27. International Union Against Cancer (UICC). TNM Classification of Malignant Tumours, Third Edition, Revised Edition. Ed. Harmer, MH: Springer, 1982.
28. International Union Against Cancer (UICC). TNM Classification of Malignant Tumours, Fourth Edition. Ed. Hermanek, P. Sobin, L.H.: Springer, 1987.
29. International Union Against Cancer (UICC). TNM Classification of Malignant Tumours, Fourth Edition, 2nd Revised Edition. Ed. Hermanek, P. Sobin, L.H.: Springer, 1992.
30. International Union Against Cancer (UICC). TNM Classification of Malignant Tumours, Fifth Edition. Ed. Hermanek, P. Sobin, L.H.: John Wiley & Sons, 1997.
31. International Union Against Cancer (UICC). TNM Classification of Malignant Tumours, Sixth Edition. Ed. Hermanek, P. Wittekind Ch: John Wiley & Sons, 2002.
32. Mardis ER. A decade's perspective on DNA sequencing technology. Nature. 2011;470: 198-203.
33. Wu X, Scelo G, Purdue MP, et al. A genome-wide association study identifies a novel susceptibility locus for renal cell carcinoma on 12p11.23. Human Molecular Genetics. 2011;21: 456-462.
34. Purdue MP, Johansson M, Zelenika D, et al. Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. Nature Genetics. 2011;43: 60-65.
35. Henrion M, Frampton M, Scelo G, et al. Common variation at 2q22.3 (ZEB2) influences the risk of renal cancer. Human Molecular Genetics. 2012;22: 825-831.
36. Scelo G, Purdue MP, Brown KM, et al. Genome-wide association study identifies multiple risk loci for renal cell carcinoma. Nature Communications. 2017;8: 15724.
37. Laskar RS, Muller DC, Li P, et al. Sex specific associations in genome wide association analysis of renal cell carcinoma. Eur J Hum Genet. 2019;27: 1589-1598.
38. Gudmundsson J, Sulem P, Gudbjartsson DF, et al. A common variant at 8q24.21 is associated with renal cell cancer. Nature Communications. 2013;4: 2776.

39. Purdue MP, Ye Y, Wang Z, et al. A Genome-Wide Association Study of Renal Cell Carcinoma among African Americans. Cancer Epidemiology Biomarkers & Prevention. 2014;23: 209.
40. Henrion MYR, Purdue MP, Scelo G, et al. Common Variation at 1q24.1 (ALDH9A1) Is a Potential Risk Factor for Renal Cancer. PLOS ONE. 2015;10: e0122589.
41. Thomas D. Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies. Annu Rev Public Health. 2010;31: 21-36.
42. Border R, Keller MC. Commentary: Fundamental problems with candidate gene-by-environment interaction studies - reflections on Moore and Thoemmes (2016). J Child Psychol Psychiatry. 2017;58: 328-330.
43. Amos W, Driscoll E, Hoffman JI. Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? Proc Biol Sci. 2011;278: 1183-1188.
44. Moore LE, Nickerson ML, Brennan P, et al. Von Hippel-Lindau (VHL) inactivation in sporadic clear cell renal cancer: associations with germline VHL polymorphisms and etiologic risk factors. PLoS Genet. 2011;7: e1002312.
45. Banks RE, Tirukonda P, Taylor C, et al. Genetic and Epigenetic Analysis of von Hippel-Lindau (VHL) Gene Alterations and Relationship with Clinical Variables in Sporadic Renal Cancer. Cancer Research. 2006;66: 2000-2011.
46. Lessi F, Mazzanti CM, Tomei S, et al. VHL and HIF-1α: gene variations and prognosis in early-stage clear cell renal cell carcinoma. Med Oncol. 2014;31: 840.
47. Wang W-C, Tsou M-H, Chen H-J, Hsu W-F, Lai Y-C. Two single nucleotide polymorphisms in the von Hippel-Lindau tumor suppressor gene in Taiwanese with renal cell carcinoma. BMC Research Notes. 2014;7: 638.
48. Lv C, Bai Z, Liu Z, Luo P, Zhang J. Renal cell carcinoma risk is associated with the interactions of APOE, VHL and MTHFR gene polymorphisms. Int J Clin Exp Pathol. 2015;8: 5781-5786.
49. Qin C, Cao Q, Ju X, et al. The polymorphisms in the VHL and HIF1A genes are associated with the prognosis but not the development of renal cell carcinoma. Annals of Oncology. 2012;23: 981-989.
50. Moore SR. Commentary: What is the case for candidate gene approaches in the era of high-throughput genomics? A response to Border and Keller (2017). Journal of child psychology and psychiatry, and allied disciplines. 2017;58: 331-334.
51. Patnala R, Clements J, Batra J. Candidate gene association studies: a comprehensive guide to useful in silico tools. BMC Genet. 2013;14: 39.
52. Thomas DC, Witte JS. Point: population stratification: a problem for case-control studies of candidate-gene associations? Cancer Epidemiol Biomarkers Prev. 2002;11: 505-512.
53. van den Oord EJ. Controlling false discoveries in candidate gene studies. Mol Psychiatry. 2005;10: 230-231.
54. Sutton ALM, Robin NH. Clinical application of whole exome sequencing: not (yet) ready for primetime. Current Opinion in Pediatrics. 2012;24.
55. Weinstein JN, Collisson EA, Mills GB, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013;45: 1113-1120.
56. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47: D941-d947.
57. Sato Y, Yoshizato T, Shiraishi Y, et al. Integrated molecular analysis of clear-cell renal cell carcinoma. Nat Genet. 2013;45: 860-867.

58. Hedegaard J, Thorsen K, Lund MK, et al. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. PLoS One. 2014;9: e98187.
59. Sarnecka AK, Nawrat D, Piwowar M, Ligęza J, Swadźba J, Wójcik P. DNA extraction from FFPE tissue samples - a comparison of three procedures. Contemporary oncology (Poznan, Poland). 2019;23: 52-58.
60. Kokkat TJ, Patel MS, McGarvey D, LiVolsi VA, Baloch ZW. Archived formalin-fixed paraffin-embedded (FFPE) blocks: A valuable underexploited resource for extraction of DNA, RNA, and protein. Biopreserv Biobank. 2013;11: 101-106.
61. Schweiger MR, Kerick M, Timmermann B, et al. Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis. PLoS One. 2009;4: e5548.
62. Smith S, Morin PA. Optimal storage conditions for highly dilute DNA samples: a role for trehalose as a preserving agent. J Forensic Sci. 2005;50: 1101-1108.
63. Ferrer I, Armstrong J, Capellari S, et al. Effects of Formalin Fixation, Paraffin Embedding, and Time of Storage on DNA Preservation in Brain Tissue: A BrainNet Europe Study. Brain Pathology. 2007;17: 297-303.
64. Ludyga N, Grünwald B, Azimzadeh O, et al. Nucleic acids from long-term preserved FFPE tissues are suitable for downstream analyses. Virchows Arch. 2012;460: 131-140.
65. Carrick DM, Mehaffey MG, Sachs MC, et al. Robustness of Next Generation Sequencing on Older Formalin-Fixed Paraffin-Embedded Tissue. PLoS One. 2015;10: e0127353.
66. Zhao Y, Mehta M, Walton A, et al. Robustness of RNA sequencing on older formalin-fixed paraffin-embedded tissue from high-grade ovarian serous adenocarcinomas. PLoS One. 2019;14: e0216050.
67. Do H, Dobrovic A. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. Clin Chem. 2015;61: 64-71.
68. de Leng WW, Gadellaa-van Hooijdonk CG, Barendregt-Smouter FA, et al. Targeted Next Generation Sequencing as a Reliable Diagnostic Assay for the Detection of Somatic Mutations in Tumours Using Minimal DNA Amounts from Formalin Fixed Paraffin Embedded Material. PLoS One. 2016;11: e0149405.
69. Smits KM, Schouten LJ, van Dijk BA, et al. Genetic and epigenetic alterations in the von hippel-lindau gene: the influence on renal cancer prognosis. Clin Cancer Res. 2008;14: 782-787.
70. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013;499: 43-49.
71. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2: 401-404.
72. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6: pl1.
73. Hiley C, de Bruin EC, McGranahan N, Swanton C. Deciphering intratumor heterogeneity and temporal acquisition of driver events to refine precision medicine. Genome Biol. 2014;15: 453.
74. Gerlinger M, Horswell S, Larkin J, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nat Genet. 2014;46: 225-233.
75. Ferronika P, Kats-Ugurlu G, Haryana SM, et al. Mutational heterogeneity between different regional tumour grades of clear cell renal cell carcinoma. Exp Mol Pathol. 2020;115: 104431.
76. Ricketts CJ, Linehan WM. Intratumoral heterogeneity in kidney cancer. Nat Genet. 2014;46: 214-215.

77. Ferronika P, Hof J, Kats-Ugurlu G, et al. Comprehensive Profiling of Primary and Metastatic ccRCC Reveals a High Homology of the Metastases to a Subregion of the Primary Tumour. Cancers (Basel). 2019;11.
78. Turajlic S, Xu H, Litchfield K, et al. Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. Cell. 2018;173: 595-610.e511.
79. Callahan CL, Hofmann JN, Corley DA, et al. Obesity and renal cell carcinoma risk by histologic subtype: A nested case-control study and meta-analysis. Cancer Epidemiol. 2018;56: 31-37.
80. Purdue MP, Moore LE, Merino MJ, et al. An investigation of risk factors for renal cell carcinoma by histologic subtype in two case-control studies. Int J Cancer. 2013;132: 2640-2647.
81. Dal Maso L, Zucchetto A, Tavani A, et al. Renal cell cancer and body size at different ages: an Italian multicenter case-control study. Am J Epidemiol. 2007;166: 582-591.
82. Colt JS, Hofmann JN, Schwartz K, et al. Antihypertensive medication use and risk of renal cell carcinoma. Cancer Causes Control. 2017;28: 289-297.
83. Gansler T, Fedewa SA, Flanders WD, Pollack LA, Siegel DA, Jemal A. Prevalence of Cigarette Smoking among Patients with Different Histologic Types of Kidney Cancer. Cancer Epidemiol Biomarkers Prev. 2020;29: 1406-1412.
84. Brannon AR, Reddy A, Seiler M, et al. Molecular Stratification of Clear Cell Renal Cell Carcinoma by Consensus Clustering Reveals Distinct Subtypes and Survival Patterns. Genes Cancer. 2010;1: 152-163.
85. Purdue MP, Rhee J, Xiaohua G, et al. Differences in Risk Factors for Molecular Subtypes of Clear Cell Renal Cell Carcinoma: a Case-Control Study. In: Proceedings of 111th Annual Meeting of the American Association for Cancer Research; 2020 June 22-24. Philadelphia (PA): AACR, 2020.
86. Linehan WM, Spellman PT, Ricketts CJ, et al. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. N Engl J Med. 2016;374: 135-145.
87. Graff RE, Sanchez A, Tobias DK, et al. Type 2 Diabetes in Relation to the Risk of Renal Cell Carcinoma Among Men and Women in Two Large Prospective Cohort Studies. Diabetes Care. 2018;41: 1432-1437.
88. Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9(th) edition. Diabetes Res Clin Pract. 2019;157: 107843.
89. Cheungpasitporn W, Thongprayoon C, O'Corragain OA, et al. The risk of kidney cancer in patients with kidney stones: a systematic review and meta-analysis. Qjm. 2015;108: 205-212.
90. Lin CL, Huang WT, Fan WC, et al. Associations between interventions for urolithiasis and urinary tract cancer among patients in Taiwan: The effect of early intervention. Medicine (Baltimore). 2016;95: e5594.
91. Sun LM, Lin CL, Chang YJ, et al. Urinary tract stone raises subsequent risk for urinary tract cancer: a population-based cohort study. BJU Int. 2013;112: 1150-1155.
92. Shuch B, Amin A, Armstrong AJ, et al. Understanding pathologic variants of renal cell carcinoma: distilling therapeutic opportunities from biologic complexity. Eur Urol. 2015;67: 85-97.
93. Schepers MSJ, Duim RAJ, Asselman M, Romijn JC, Schröder FH, Verkoelen CF. Internalization of calcium oxalate crystals by renal tubular cells: A nephron segment–specific process? Kidney International. 2003;64: 493-500.
94. Khan SR. Renal tubular damage/dysfunction: key to the formation of kidney stones. Urological Research. 2006;34: 86-91.
95. Deckers IA, van den Brandt PA, van Engeland M, et al. Long-term dietary sodium, potassium and fluid intake; exploring potential novel risk factors for renal cell cancer in the Netherlands Cohort

Study on diet and cancer. Br J Cancer. 2014;110: 797-801.
96. Antwi SO, Eckel-Passow JE, Diehl ND, et al. Alcohol consumption, variability in alcohol dehydrogenase genes and risk of renal cell carcinoma. Int J Cancer. 2018;142: 747-756.
97. Deckers IA, van den Brandt PA, van Engeland M, et al. Polymorphisms in genes of the renin-angiotensin-aldosterone system and renal cell cancer risk: interplay with hypertension and intakes of sodium, potassium and fluid. Int J Cancer. 2015;136: 1104-1116.
98. Karami S, Brennan P, Navratilova M, et al. Vitamin d pathway genes, diet, and risk of renal cell carcinoma. Int J Endocrinol. 2010;2010: 879362.
99. Smits KM, Schouten LJ, van Dijk BA, et al. Polymorphisms in genes related to activation or detoxification of carcinogens might interact with smoking to increase renal cancer risk: results from The Netherlands Cohort Study on diet and cancer. World J Urol. 2008;26: 103-110.
100. Melkonian SC, Daniel CR, Ye Y, et al. Gene-environment interaction of genome-wide association study-identified susceptibility loci and meat-cooking mutagens in the etiology of renal cell carcinoma. Cancer. 2016;122: 108-115.
101. Hakimi AA, Mano R, Ciriello G, et al. Impact of recurrent copy number alterations and cancer gene mutations on the predictive accuracy of prognostic models in clear cell renal cell carcinoma. J Urol. 2014;192: 24-29.
102. Patard JJ, Fergelot P, Karakiewicz PI, et al. Low CAIX expression and absence of VHL gene mutation are associated with tumor aggressiveness and poor survival of clear cell renal cell carcinoma. Int J Cancer. 2008;123: 395-400.
103. Voss MH, Reising A, Cheng Y, et al. Genomically annotated risk model for advanced renal-cell carcinoma: a retrospective cohort study. Lancet Oncol. 2018;19: 1688-1698.
104. Hakimi AA, Chen YB, Wren J, et al. Clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma. Eur Urol. 2013;63: 848-854.
105. Hakimi AA, Ostrovnaya I, Reva B, et al. Adverse outcomes in clear cell renal cell carcinoma with mutations of 3p21 epigenetic regulators BAP1 and SETD2: a report by MSKCC and the KIRC TCGA research network. Clin Cancer Res. 2013;19: 3259-3267.
106. Grossman RL, Heath AP, Ferretti V, et al. Toward a Shared Vision for Cancer Genomic Data. N Engl J Med. 2016;375: 1109-1112.
107. Wang X, Steensma JT, Bailey MH, Feng Q, Padda H, Johnson KJ. Characteristics of The Cancer Genome Atlas cases relative to U.S. general population cancer cases. Br J Cancer. 2018;119: 885-892.
108. Büscheck F, Fraune C, Simon R, et al. Prevalence and clinical significance of VHL mutations and 3p25 deletions in renal tumor subtypes. Oncotarget. 2020;11: 237-249.
109. Sharifi N, Farrar WL. Perturbations in hypoxia detection: a shared link between hereditary and sporadic tumor formation? Med Hypotheses. 2006;66: 732-735.
110. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578: 94-101.
111. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. Nature. 2013;500: 415-421.
112. Begg CB, Seshan VE, Zabor EC, et al. Genomic investigation of etiologic heterogeneity: methodologic challenges. BMC Medical Research Methodology. 2014;14: 138.

ADDENDUM

Summary

The aetiology of RCC is still poorly understood. Although multiple risk factors have been investigated in relation to the risk of renal cell carcinoma (RCC), few risk factors have consistently been associated with RCC. In addition, there has been little attention for potential heterogeneity of associations across RCC subtypes. In the past, studies primarily focused on overall RCC. In recent years more attention has been given to associations between risk factors and distinct histological subtypes of RCC. Due to their distinct clinical, pathological and genetic makeup, it has been hypothesized that these histological subtypes may be differentially associated with risk factors. In turn, these differences may also contribute to a difference in aetiology across histological subtypes. Overall, more studies are needed on risk factors to unravel these relationships and to get a clearer view on the aetiology of RCC and its most common histological subtypes clear-cell RCC (ccRCC) and papillary RCC (pRCC).

All studies within this thesis were conducted using information obtained through the Netherlands Cohort Study on diet and cancer (NLCS). The NLCS was initiated in September 1986 with the inclusion of 120,852 men and women aged 55-69 years. Information on dietary habits and other risk factors for cancer, such as lifestyle factors, medical conditions and anthropometry, were collected through a mailed, self-administered questionnaire. In addition to the questionnaire, approximately 90,000 participants provided toenail clippings. Follow-up for cancer occurrence was performed for all participants through computerized record linkage with the Netherlands Cancer Registry (NCR), the Dutch pathology registry (PALGA), and the causes of death registry maintained by Statistics Netherlands (CBS). During 20.3 years of follow-up 608 RCC cases were identified. For 454 RCC cases, formalin-fixed paraffin-embedded (FFPE) tumour tissue could be collected and was available for the revision of tumour histology by two experienced pathologists. Of these 454 RCC cases, 366 were classified as ccRCC cases, 60 as pRCC cases, and 28 as other or undefined RCC cases. DNA Isolated from FFPE healthy tissue and from provided toenail clippings was used for genotyping germline single nucleotide variants. In addition, DNA was isolated from collected FFPE tumour blocks for the identification of somatic mutations.

In **chapter 2** of this thesis, we focused on the relationship between type 2 diabetes mellitus and the risk of RCC. We observed that type 2 diabetes was moderately associated with an increased RCC risk. In particular, we observed that female participants with type 2 diabetes had a more strongly elevated RCC risk, while this association was not present in men. Furthermore, we investigated the use of anti-diabetic medication and RCC risk. We observed that individuals with type 2 diabetes who reported the use of insulin had an increased risk of RCC, when compared to participants without type 2 diabetes. However, the number of people in this particular analysis was quite small, which might influence the accuracy of this result.

In **chapter 3** we analysed whether there was a difference in association between established modifiable risk factors for RCC and the development of the two most common histological subtypes of RCC. To this end, we focused on the heterogeneity between body mass index (BMI), cigarette smoking, alcohol consumption and hypertension across ccRCC and pRCC risk. We observed a potential aetiologic heterogeneity regarding the association for BMI. BMI was positively associated with the risk of ccRCC, but inversely with pRCC risk. This

finding was of great interest as BMI has been indicated as a potential source of aetiologic heterogeneity in prior studies on RCC subtypes. Furthermore, no clear and consistent indications were found for other sources of heterogeneity of associations regarding the risk of RCC subtypes.

In the analyses presented in **chapter 4**, we observed a relationship between the self-reported history of kidney stones and the development of RCC and upper tract urothelial carcinoma (UTUC). In particular, the risk of kidney stones was associated with an increased risk of papillary RCC, but not ccRCC. No heterogeneity of associations was found between kidney stones and UTUC in the ureter and renal pelvis. These findings raise questions regarding the mechanisms by which kidney stones may increase the risk of RCC, as the general hypothesis that kidney stones cause chronic inflammation and infection which may lead to the development of a tumour would not explain the observed differences in risk between ccRCC and pRCC.

Genetic alterations in the VHL/HIF pathway have been found to be important drivers of carcinogenesis in ccRCC. Therefore, we investigated the role of three candidate Single Nucleotide Polymorphisms (SNPs) in *VHL* and one SNP in *HIF1A* on the risk of developing (cc)RCC, their potential for interplay with the environment and with each other, and the relationship between *VHL* SNPs and *VHL* promoter methylation (**chapter 5**). One *VHL* SNP, VHL_rs779805, was associated with an increased RCC risk. No associations were observed for the three other SNPs. In addition, no clear interactions were found between the four selected SNPs and environmental factors after adjusting for multiple testing. Neither did we observe interactions between the SNPs. Lastly, we observed no associations between the selected *VHL* SNPs and *VHL* promoter methylation.

For the analyses presented in **chapter 6** of this thesis we performed targeted sequencing on DNA isolated from tumour blocks of 252 ccRCC cases. Using this sequencing information, we created a seven-gene mutational profile, based on the seven genes with the highest reported mutation frequencies in ccRCC, to identify somatic alterations that alter the prognosis of ccRCC. This panel included *VHL, PBRM1, SETD2, BAP1, MTOR, KDM5C,* and *TP53*. Overall, 110 cases were eligible for these analyses. Individuals with mutations in *VHL* and *PBRM1* had a more favourable ccRCC-specific survival, compared to individuals without mutations in these genes. However, this association did not maintain statistical significance after correction for multiple testing. The reasons for this potential association remain poorly understood. However, both *VHL* and *PBRM1* have been implicated as important drivers in the early events of ccRCC development.

In conclusion, taking into account various methodological considerations detailed in **chapter 7**, this thesis highlights the importance of further subclassification of renal cancers by histopathological features and molecular characteristics for both aetiology and progression. This subclassification, however, does not come without compromise and will require great adaptability by epidemiological researchers. In the future, large-scale efforts will be required to be able to accurately and consistently assess the role of risk factors on the aetiology of histologic subtypes of RCC.

Nederlandse Samenvatting

Nog steeds is er weinig inzicht in de etiologie van het niercelcarcinoom (RCC). Hoewel verscheidene risicofactoren onderzocht zijn in relatie tot het risico op RCC, zijn er nog altijd weinig risicofactoren consistent in verband gebracht met RCC. Bovendien is er tot op heden weinig aandacht voor de mogelijke heterogeniteit van associaties met de verschillende RCC-subtypen. In het verleden waren onderzoeken voornamelijk gericht op RCC in zijn geheel. De laatste jaren is echter meer aandacht besteed aan de specifieke associaties tussen risicofactoren en verschillende histologische subtypen van RCC. Verondersteld wordt dat deze subtypen mogelijk verschillend geassocieerd kunnen zijn met diverse risicofactoren vanwege de verschillende klinische, pathologische en genetische samenstelling. Deze verschillen kunnen op hun beurt ook bijdragen aan een verschil in oorzaken tussen histologische subtypen. Er is meer data nodig om deze relaties te ontrafelen en om een duidelijker beeld te krijgen van de etiologie van het RCC en de meest voorkomende histologische subtypen van het RCC, namelijk heldercellig RCC (ccRCC) en papillair RCC (pRCC).

De onderzoeken in dit proefschrift zijn allemaal uitgevoerd met behulp van informatie verkregen uit de Nederlandse Cohortstudie naar voeding en kanker (NLCS). De NLCS is in september 1986 gestart met de inclusie van 120.852 mannen en vrouwen die op dat moment 55-69 jaar oud waren. Door middel van een zelf in te vullen vragenlijst werd informatie verzameld over voedingsgewoonten en andere risicofactoren voor kanker, waaronder leefstijlfactoren, medische aandoeningen en antropometrie. Daarnaast leverden ongeveer 90.000 deelnemers teennagelknipsels aan. Met behulp van een koppeling met de Nederlandse Kankerregistratie (NKR), het Nederlandse pathologieregister (PALGA) en de registratie van doodsoorzaken, zoals bijgehouden door het Centraal Bureau voor de Statistiek (CBS), werd geregistreerd welke deelnemers, gedurende de studieperiode van meer dan 20 jaar, nierkanker ontwikkelden. In totaal werden gedurende de studieperiode 608 RCC-gevallen geïdentificeerd. Van454 RCC-gevallen kon formaline-gefixeerd paraffine-ingebed (FFPE) tumorweefsel worden verzameld. Dit FFPE-weefsel kon gebruikt worden voor de herziening van de tumorhistologie (subclassificatie) door twee ervaren uropathologen. Van deze 454 RCC-gevallen werden 366 RCC-gevallen geclassificeerd als ccRCC, 60 als pRCC en 28 als andere of niet-gedefinieerde RCC. Daarnaast werd DNA geïsoleerd uit gezond FFPE-weefsel en uit de verstrekte teennagelknipsels voor het genotyperen van kiembaan mutaties en werd DNA geïsoleerd uit verzamelde FFPE-tumorblokken voor de identificatie van somatische mutaties.

In **hoofdstuk 2** van dit proefschrift hebben we gefocust op de relatie tussen diabetes mellitus type 2 en het risico op RCC. We zagen dat diabetes type 2 geassocieerd was met een licht verhoogd risico op het krijgen van RCC. We zagen in het bijzonder dat vrouwelijke deelnemers met diabetes type 2 een sterker verhoogd RCC-risico hadden, terwijl deze associatie niet aanwezig was bij mannen. Verder onderzochten we het gebruik van antidiabetica en het RCC-risico. We stelden vast dat personen met diabetes type 2 die aangaven dat ze insuline gebruikten een verhoogd risico hadden op het krijgen van RCC, in vergelijking met deelnemers zonder diabetes type 2. Het aantal mensen in deze specifieke analyse was echter vrij klein, waardoor de nauwkeurigheid van dit resultaat beïnvloedt zou kunnen zijn.

In **hoofdstuk 3** hebben we onderzocht of er een verschil was in het verband tussen bekende beïnvloedbare risicofactoren voor RCC en de ontwikkeling van de twee meest voorkomende histologische subtypes van RCC. Daartoe hebben we ons gericht op de heterogeniteit tussen body mass index (BMI), het roken van sigaretten, alcoholgebruik en hypertensie op het ccRCC- en pRCC-risico. We constateerden een mogelijke etiologische heterogeniteit met de associatie voor BMI. BMI was namelijk geassocieerd met een verhoogd risico op ccRCC, maar omgekeerd geassocieerd met het risico op het krijgen van pRCC. Deze bevinding kwam overeen met eerdere onderzoeken naar RCC-subtypen, waarin BMI ook als mogelijke bron van etiologische heterogeniteit werd geduid. Voor andere risicofactoren werden geen duidelijke of consistente aanwijzingen gevonden voor heterogeniteit met betrekking tot het risico van RCC-subtypen.

In de analyses die beschreven staan in **hoofdstuk 4** hebben we een relatie gevonden tussen de zelf-gerapporteerde geschiedenis van nierstenen en de ontwikkeling van RCC en urotheelcarcinoom van de bovenste urinewegen (UTUC). In het bijzonder was het risico op nierstenen geassocieerd met een verhoogd risico op pRCC, maar niet op het ccRCC-risico. Tussen nierstenen en UTUC in de urineleider en het nierbekken werd geen heterogeniteit van associaties gevonden. Deze bevindingen roepen vragen op over het onderliggende mechanisme waarmee nierstenen het risico op RCC verhogen. De algemene hypothese die beschrijft dat nierstenen chronische ontstekingen en infecties veroorzaken die vervolgens kunnen leiden tot de ontwikkeling van een tumor, zou vanwege deze bevindingen geen passende verklaring zijn voor de waargenomen verschillen in het risico tussen ccRCC en pRCC.

Op basis van eerder gepubliceerde artikelen blijken genetische veranderingen in het *VHL/HIF*-systeem belangrijke beïnvloedende factoren van carcinogenese bij ccRCC te zijn. Daarom onderzochten we de rol van drie veelvoorkomende enkelvoudige basenpaarveranderingen (single nucleotide polymorphisms; SNP's) in *VHL* en één SNP in *HIF1A* op het risico van het ontwikkelen van (cc)RCC **(hoofdstuk 5)**. Daarnaast onderzochten we de interactie van deze SNPs met verscheidene omgevingsfactoren en met elkaar. Tevens onderzochten we de relatie tussen de *VHL* SNPs en *VHL* promotor methylering. Eén *VHL* SNP, namelijk *VHL_*rs779805, was geassocieerd met een verhoogd (cc)RCC-risico. Er werden geen associaties waargenomen voor de drie andere SNP's. Bovendien werden geen duidelijke interacties gevonden tussen de vier geselecteerde SNP's en omgevingsfactoren na het uitvoeren van een correctie voor meervoudig toetsen. Evenmin hebben we interacties tussen de SNP's waargenomen. Ten slotte, namen we ook geen associaties waar tussen de geïncludeerde *VHL* SNP's en *VHL*-promotormethylering.

Voor de analyses in **hoofdstuk 6** van dit proefschrift hebben we een gerichte nucleotide volgorde bepaling (targeted sequencing) uitgevoerd op DNA geïsoleerd uit tumorblokken van 252 ccRCC gevallen. Met behulp van deze sequentie informatie hebben we een mutatieprofiel gemaakt, welke gebaseerd is op de zeven genen (*VHL, PBRM1, SETD2, BAP1, MTOR, KDM5C* en *TP53)* met de hoogst gerapporteerde mutatiefrequenties in ccRCC. Op deze manier wilden wij somatische veranderingen identificeren die mogelijk de prognose van ccRCC beïnvloeden. In totaal kwamen 110 cases in aanmerking voor onze analyses. Uit de analyses bleek dat cases met mutaties in *VHL* en *PBRM1* een gunstiger ccRCC-specifieke

overleving hadden in vergelijking met cases zonder mutaties in deze genen. Deze associatie verloor echter de statistische significantie na correctie voor meervoudig toetsen. De redenen voor het verband van deze genen met een verbeterde ccRCC-overleving blijft onduidelijk. Uit eerdere studies is echter gebleken dat zowel *VHL* als *PBRM1* betrokken zijn bij de initiële ccRCC-ontwikkeling. Verder onderzoek is nodig om te zien of deze genen ook een rol spelen bij de prognose van ccRCC.

Rekening houdend met verschillende methodologische overwegingen, zoals beschreven in **hoofdstuk 7**, benadrukt dit proefschrift het belang van verdere subclassificatie van nierkanker op basis van histopathologische kenmerken en moleculaire karakteristieken voor zowel de etiologie als de progressie van nierkanker. Verdere subclassificatie komt echter niet zonder compromis en vereist een groot aanpassingsvermogen van epidemiologische onderzoekers. Naar verwachting zal in de toekomst grootschalig onderzoek nodig zijn om de rol van risicofactoren op de etiologie van histologische subtypen van RCC nauwkeurig en consistent te kunnen beoordelen.

Impact paragraph

In this chapter I will discuss the scientific and social impact that the research described in this dissertation has provided in the short-term and could provide in the long-term. Furthermore, we will detail how the results were disseminated during the PhD trajectory. Additionally, we put the results of this thesis into a broader perspective by highlighting the potential impact of this dissertation for researchers, clinicians and policymakers in a public health setting and patient care.

The dissemination of results during the PhD trajectory

During the PhD trajectory various means were employed to disseminate the study results to a broad audience, including researchers and clinicians from multiple disciplines.

Firstly, study findings were published in various international scientific journals, as detailed in the beginning of each chapter of this dissertation. Secondly, our scientific findings were presented at various conferences and symposia for audiences with a broad background. For example, in 2017 and 2019, our scientific results were presented at the Dutch Epidemiological Conference (WEON), which is generally attended by epidemiologists from various research backgrounds. In addition, our research was presented at science days and research meetings in Maastricht thereby sharing the results to multidisciplinary audiences with a broad clinical and/or healthcare background. Thirdly, scientific results were presented targeted to specific audiences relevant to the scientific work. For instance, considerations regarding the feasibility of the use of formalin-fixed paraffin embedded (FFPE) tumour material were presented at the 4th International Molecular Pathological Epidemiology (MPE) Meeting (Boston, USA) in 2018. At this conference, an international audience from diverse fields gathered for discussions on the topic of MPE. This event proved to be a key opportunity to discuss our insights into factors affecting the quality of sequencing when using routinely archived FFPE material with experts in the field of DNA sequencing. During this meeting, we realized that the use of FFPE tissue for sequencing wasn't as clear-cut as we expected and that various research groups were experiencing similar difficulties in maintaining a high sequencing quality needed in research. By sharing our first-hand experiences from the Netherlands Cohort Study on diet and cancer (NLCS), we hopefully gave other researchers insights on what to account for when using routinely archived FFPE tumour material that has been stored for extended periods of time. Hopefully, this may lead to a reduction in research waste, as researchers may be more inclined to account for sample characteristics (*i.e.* storage duration and DNA concentration) during DNA isolation and work-up that could lead to reduced sequencing yields. To further the dissemination of these insights, appendix 2 has been added to **chapter 5** in this dissertation. Moreover, factors influencing the quality of sequencing were presented to pathologists at the Maastricht Pathology Meeting (Maastricht, the Netherlands) in 2018, which was a joint effort from the British Division of the International Academy of Pathology (BDIAP), the Pathological Society of Great Britain & Ireland and the Dutch Society for Pathology (NVVP). During this meeting questions were addressed on what factors pathologists, among others, should be aware of to maintain sufficient data quality for sequencing. Sharing these insights could be useful for expectations management in the case researchers want to use routinely archived FFPE tumour material which has been stored in suboptimal conditions for a prolonged period of time. Lastly, these insights could also provide helpful information for the initiation of future projects to enable researchers to

optimize the collection and storage of (FFPE) tumour tissue.
Lastly, our research on the association between kidney stones and renal cell carcinoma (RCC) and upper tract urothelial carcinoma (UTUC) was picked up by international press, which enabled our research to reach a broad audience beyond the scientific community. Furthermore, the article was featured on online websites specialized in clinical research, which may hopefully have led to additional awareness around the potential link between kidney stones and RCC and UTUC and its subtypes in clinicians and other health care professionals.

Future impact of the generated knowledge
The findings presented in this dissertation may also have various scientific and social implications in the (nearby) future. In addition, several challenges for research are brought up by the research presented in this dissertation that researchers should be aware of in the field of RCC research.

Determinants for renal cell carcinoma
One of the primary aims of this dissertation was to assess risk factors for RCC and its subtypes. Two potential risk factors which are highly prevalent in the population, namely type 2 diabetes mellitus and kidney stones, which as of yet are not considered established risk factors for RCC, were observed to be associated with an increased risk of RCC in the Netherlands Cohort Study on diet and cancer.
The study on type 2 diabetes mellitus described in this dissertation reinforces recent evidence from the Nurses' Health Study and the Health Professionals Follow Up Study that diabetes mellitus is associated with RCC specifically in women, but not in men[1]. From a scientific perspective, this observation could be of great importance for finding clues for unravelling factors that contribute to the development of RCC. In a similar vein, the observed association between anti-diabetic medication use and the risk of RCC may prove to be an interesting observation for future research. Anti-diabetic medication has often been a source for heated debate in the scientific community. This is, in part, due to the inconsistent associations between anti-diabetic medication and the risk of various cancers[2, 3]. The information regarding the effect of these types of medication on the risk of kidney cancer remains limited from observational studies[2]. Hopefully, other largescale prospective cohorts will find opportunities to look into replication of these findings, as these observed results in our study do not yet suffice as conclusive evidence due to underlying methodological constraints.
The current evidence on the association between kidney stones and RCC is slightly stronger, although mainly supported by evidence from retrospective cohort studies or case-control studies. As a result of a meta-analysis by Cheungpasitporn *et al.*, kidney stones have been featured on Wolters Kluwer UpToDate as a risk factor for RCC, enabling physicians to easily access information on the role of this risk factor in RCC[4, 5]. Interestingly, the majority of studies featured in the aforementioned meta-analysis were published between 1984-1997[4]. Resultingly, no evidence is available regarding associations between kidney stones and specific RCC tumor entities. Therefore, if clinicians are inclined to look for more information on kidney stones and the relationship with RCC subtypes, they are likely to end up with the data from our study, being the first to report on the heterogeneity of associations regarding kidney stones across ccRCC and pRCC risk. This heterogeneity of associations is of particular interest, as the relationship cannot be explained by the occurrence of kidney stones due to a

disconnect between the location where kidney stones form and the location where ccRCC and pRCC develop. This finding may therefore provide a lead for delving deeper into the characteristics and mechanisms resulting in the association between kidney stones and RCC subtypes for future research.
An increased awareness by both clinicians, policy makers and patients for (lesser known) risk factors for RCC and its subtypes may contribute to improved individual healthcare in the future. With increased awareness clinicians may be better equipped to provide advice targeted to the characteristics of the patient. As many of the risk factors for RCC are shared with other important comorbidities that tend to severely effect the quality of life, these advices may improve the overall health of patients. Furthermore, policy makers and health professionals may include these risk factors in guidelines and factsheets to aid in the education of clinicians, patients and researchers.

Heterogeneous associations for histologic subtypes of RCC
In recent years more attention has been brought to the associations between risk factors and the risk of specific histological subtypes of RCC. These histological subtypes may be differentially associated to risk factors due to their distinct clinical, pathological and genetic makeup. Information on the association between risk factors and specific subtypes of RCC is crucial as potential differences in association can add noise when trying to assess aetiological relationships. For instance, if a risk factor is only associated to a specific histologic subtype of RCC, no clear or conclusive evidence may be found if studies are performed in a population with varying tumour histologic subtypes. Therefore, the observed differences across the risk of ccRCC and pRCC regarding the history of kidney stones and body mass index (BMI) may help elucidate differences in aetiology and differences in aetiological mechanisms involved between subtypes. At present, new entities are continuously described and incorporated in the WHO classification for tumours. Therefore, more research is needed to establish whether different entities possess different aetiologic mechanisms.

Even though established risk factors convey additional risks for RCC it remains hard to translate these findings into direct changes in prevention strategies for health policy makers. Population screening, for instance, is not a feasible strategy at this point in time. One of the primary barriers to population screening for RCC is the relatively low prevalence of RCC. It is estimated that by screening 1,000 asymptomatic individuals from the general population using ultrasound only one or two cases of RCC will be detected[6]. An employable strategy to resolve this problem is by creating prediction models for targeted screening of high-risk populations. However, as (modifiable) risk factors for RCC, including smoking, obesity, hypertension, kidney stones and diabetes mellitus, only translate into moderate risk increases and since they are not specific for RCC, these factors are unlikely to be suitable for use in risk prediction models[6]. Resultingly, more evidence is needed on the aetiology of RCC to make risk estimations more accurate. Hopefully, this will enable us to detect RCC in earlier stages in future times and, in turn, reduce mortality rates of RCC.

For clinicians and general practitioners it remains of great importance to actively promote healthy lifestyles as the reduction of the prevalence of obesity, smoking and hypertension in the general population may pose viable strategies to the reduction of the burden of RCC. For instance, information from GLOBOCAN (Global Cancer Incidence, Mortality and

Prevalence) indicates that more than 20% of the RCC cases in Europe can be attributed to obesity[7]. Furthermore, tobacco smoking also heavily contributes to the burden of kidney cancer[8]. More information is needed to get more accurate estimations on the burden of RCC attributed to hypertension, diabetes mellitus and kidney stones. To this end, more information is needed on the direct mechanisms by which these risk factors are associated to RCC. An example is inconclusive evidence regarding the role of anti-hypertensive medication on the risk of RCC. Insights in such factors may then open new avenues for prevention strategies against renal cancer in the general population.

In the past, strategies have been made to create risk profiles of patients who developed (cc) RCC[9]. In general, the focus in these methods relies on clinical characteristics to create a prognostic index. Based on the predicted prognosis of these patients, different systemic therapies are recommended. These models could also benefit from additional prognostic variables such as molecular testing. For instance, in this dissertation **(chapter 6)** we reported that patients with ccRCC had a better prognosis if they had *VHL* or *PBRM1* mutations, when compared to patients without these mutations. Insights like these, combined with detailed clinical information may be used in the future to predict the prognosis of (cc)RCC in patients or to recommend treatments with higher rates of success.
One important sidenote regarding this observed association is that the analyses in this dissertation were based on a mostly treatment-naïve population. This may provide indications of how ccRCC progresses without the effect of interventions beyond surgical resection. Nowadays, this information is hard to obtain, as patients tend to receive various forms of treatment (a.o surgical resection, targeted therapy and/or immunotherapy). Hopefully, the information on gene profiles described in this thesis may provide useful information for researchers and clinicians who are looking to further incorporate mutational profiles in prognostic indices. Hopefully, in the future extensive tailored treatment strategies will become available for the treatment of RCC depending on the genotypic and phenotypic features of the tumour. Having clearer information regarding more optimal treatment strategies for patients by incorporating prognostic features may aid clinicians in providing better patient care in the future, which may reduce the health care burden of RCC.

Conclusion
In this dissertation we have included several studies that highlight aetiological and prognostic mechanisms for the risk of RCC and its subtypes. To disseminate this evidence to the right audiences and to increase the impact of this work, several strategies were employed to share these results with researchers, clinicians and general audiences. While most results are not directly suited for inclusion in adjustments to current healthcare, they may provide crucial insights and leads for future studies into the aetiology of RCC and its subtypes.

References

1. Graff RE, Sanchez A, Tobias DK, et al. Type 2 Diabetes in Relation to the Risk of Renal Cell Carcinoma Among Men and Women in Two Large Prospective Cohort Studies. Diabetes Care. 2018;41: 1432-1437.
2. Karlstad O, Starup-Linde J, Vestergaard P, et al. Use of insulin and insulin analogs and risk of cancer - systematic review and meta-analysis of observational studies. Curr Drug Saf. 2013;8: 333-348.
3. Zhang K, Bai P, Dai H, Deng Z. Metformin and risk of cancer among patients with type 2 diabetes mellitus: A systematic review and meta-analysis. Prim Care Diabetes. 2020.
4. Cheungpasitporn W, Thongprayoon C, O'Corragain OA, et al. The risk of kidney cancer in patients with kidney stones: a systematic review and meta-analysis. Qjm. 2015;108: 205-212.
5. Atkins MB, Choueiri TK. Epidemiology, pathology, and pathogenesis of renal cell carcinoma. In: Richie JP, editor. UpToDate. Waltham, MA, 2020.
6. Rossi SH, Klatte T, Usher-Smith J, Stewart GD. Epidemiology and screening for renal cancer. World J Urol. 2018;36: 1341-1353.
7. Arnold M, Lam F, Ervik M, Soerjomataram I. Cancer and Obesity: Global burden of cancer attributable to excess weight. Available from URL: http://gco.iarc.fr/obesity [accessed November 25, 2020].
8. Safiri S, Kolahi A-A, Mansournia MA, et al. The burden of kidney cancer and its attributable risk factors in 195 countries and territories, 1990–2017. Scientific Reports. 2020;10: 13862.
9. Achermann C, Stenner F, Rothschild SI. Treatment, Outcome and Prognostic Factors in Renal Cell Carcinoma - A Single Center Study (2000-2010). J Cancer. 2016;7: 921-927.

Dankwoord

Gedurende de afgelopen vier jaren heb ik veel verschillende mensen mogen ontmoeten en met veel geweldige mensen samen mogen werken. Door jullie allen heb ik gedurende de afgelopen jaren een behoorlijke groei op professioneel als persoonlijk vlak door kunnen en mogen maken. In dit hoofdstuk wil ik graag enkele personen of groepen even uitlichten om in het bijzonder te bedanken.

Allereerst wil ik natuurlijk mijn promotieteam bedanken.
Leo, ik heb ontzettend genoten van onze samenwerking. Ik heb veel geleerd van de kalme en weldoordachte wijze waarmee jij (epidemiologische) vraagstukken aanpakt. Bij jou kon ik niet alleen altijd terecht voor een technische of vakinhoudelijke vraag, maar ook voor het over en weer kaatsen van ideeën. Gelukkig hebben we gedurende de afgelopen jaren geleerd om al deze creativiteit te bundelen en om zo (enigszins) gestructureerd onze vraagstukken aan te pakken. Dat was tijdens onze eerste overleggen wel anders. Uiteindelijk denk ik dat we wel 3 á 4 boekjes met ideeën hadden kunnen vullen... Verder wil ik je ook bedanken voor alle kansen en mogelijkheden die je me geboden hebt. Voor mijn gevoel heb ik bij jou vanaf mijn eerste dag als PhD alle kansen gekregen die jij me kon bieden. Door jouw aanpak en goede begeleiding ben ik de afgelopen jaren niet alleen flink gegroeid als epidemioloog, maar ook als persoon.

Piet, bedankt het vertrouwen in mij. Jouw scherpe blik en kritische vragen gaven vaak een nieuwe impuls aan ons onderzoek. Ook jouw tijdsmanagement bleek van groot belang gedurende mijn promotietraject waar jij op de juiste momenten aan de rem durfde te trekken, danwel randzaken durfde weg te halen. Tevens wil ik je bedanken voor de kans om mijn onderzoek uit te voeren binnen de NLCS. Zonder jou was dit niet mogelijk geweest.

Klaas, jij hebt mij veel inzicht geboden in de wereld van genetica, sequencing en labwerk, waar ik als relatieve leek instapte aan het begin van dit promotietraject. Met jouw Groningse nuchterheid wist je me de belangrijkste concepten van genetica en sequencing in rap tempo helder uit te leggen, zodat we snel theorie en praktijk konden combineren binnen onze samenwerking. Zowel tijdens mijn bezoekjes aan Groningen, als tijdens het schrijven van de stukken voor dit boekje, maakte je altijd tijd vrij om te zorgen dat alles in orde kwam.

Verder zou ik ook graag mijn dank uit willen spreken voor de leden van de beoordelingscommissie: Prof. dr. Jeroen Kooman, Prof. dr. Axel zur Hausen, Prof. dr. Bart Kiemeney, Prof. dr. Peter Mulders, en Prof. dr. Frans Ramaekers voor de tijd en moeite die jullie gestoken hebben in het beoordelen van dit proefschrift.

Ook wil ik alle co-auteurs bedanken voor jullie waardevolle inzichten, kritische remarks en vooral de prettige samenwerking. In particular, I would like to mention Ferronika, thank you for your effort on the Maastricht/Groningen project (depending where you're doing your PhD!) and thank you for the good times when I was visiting Groningen. I am glad we were able to include our interesting paper in our dissertations.

Ik wil ook het nieroverleg speciaal benoemen dat samen met Sophie opnieuw in het leven

geroepen werd. Het informele karakter en het overmatige enthousiasme tijdens de overleggen, presentaties en het delen van hersenspinsels was altijd een ideale broedvijver voor nieuwe ideeën, artikelen en samenwerkingen. In het bijzonder wil ik Manon, Kim en Maureen, als aanschuivers van het eerste uur, bedanken voor de frisse en reeële blik. Andere Kim (haha), bedankt dat jij het aandurfde om het stokje van mij over te nemen. Dit is met alle schema's, annuleringen, etcetera, een behoorlijk uitdagende taak. Adriaan, bedankt voor de peptalk op het PathSoc en MPE congres. Selena en Iryna, hopelijk kunnen we in de toekomst nog veel mooie inzichten halen uit het revisie-project! Jaleesa... zelfs als ik paniekerig aangestuifd kwam, wist jij de kalmte te bewaren. Bedankt voor al je werk met de GENRE series en het overzicht dat je meermaals kon bieden.

Ik wil ook alle collega's van de NLCS bedanken. Middels jullie inzichten en kennis heb ik ontzettend veel geleerd. Al jullie projecten, ideeën en bezigheden blijven mij intrigeren en laten mij duizelen over alle mogelijkheden.

Verder wil ik ook alle collega's van de vakgroep epidemiologie bedanken. Jullie hebben er voor gezorgd dat ik net iets vaker koffie wilde halen, zodat ik de kans had om iemand tegen te komen voor een praatje bij het koffiezetapparaat of bij een toevallig openstaande, danwel openslaande deur. Bedankt voor jullie collegialiteit, ondersteuning en belangstelling die ik heb ervaren in alle jaren dat ik op de vakgroep rondzwierf. Ook wil ik mijn waardering uitspreken voor jullie inzet voor het onderwijs. Voor jullie studenten gaan jullie allemaal door het vuur.
Ook al was ik niet vaak zo'n binnenloper, toch wil ik graag in het bijzonder ook Yvonne Leenders, Mariëlle, Petra, Conny, Irma en Yvonne (van den Bergh) bedanken voor al jullie hulp als ik dan toch een prangende vraag had of ondersteuning vanuit de vakgroep nodig had. Jolanda, bedankt voor het inplannen van mijn niet-altijd-driemaandelijkse afspraken met Piet en Leo. Ook wil ik je bedanken voor de geweldige verjaardags-plakaten die je gemaakt hebt in de afgelopen jaren (zowel voor mij als voor collega's). Zelfs al was het onderwerp voor jou wereldvreemd, jij kwam altijd aanzetten met schitterende, doch relevante kunstwerken. Harry en Jos, jullie stonden altijd paraat om mij uit de brand te helpen. Jos, jouw aanpakkersmentaliteit en enorme collegialiteit heeft me ontzettend geholpen in de laatste jaren. Van een spontaan verdwenen beeldscherm tot een trage computer, jij wist altijd met een passende oplossing te komen. Harry, de rust en kalmte waarmee jij problemen oplost, zelfs als ik niet helemaal begreep wat er uiteindelijk aan de hand was, is van niet te onderschatten waarde. En als laatste Matty, bedankt dat je altijd een luisterend oor bood en naar mogelijkheden zocht om net iets meer te kunnen bieden. Daarnaast ontzettend bedankt voor al je hulp in de afgelopen periode, het heeft mij veel stress gescheeld wetende dat ik op jou kon vertrouwen.

Christel en Martien, ik herinner me de dag dat ik kwam solliciteren voor de HSRM nog als gisteren. Christel, zoals altijd, kwebbelend over van alles en nog wat en Martien pijnzend over de materie. Dat beeld is zeker blijven hangen gedurende de laatste jaren. Christel, bedankt voor alle momenten dat ik bij jou binnen kon lopen of dat je jezelf naar de verste uithoek van Deb bewoog. Voor je verhuizing op Deb had ik altijd een goed excuus om naar de 'Gym' te gaan, al kon dat zelfs spontaan veranderen in een korte excursie rond het gebouw. Martien, bedankt dat je altijd openstond voor vragen over (onder andere) het onderwijs en dat

je voor iedereen de tijd nam. Wat heb ik genoten van de epistels die je op de mail zette over de taken, zodat ik altijd goed voorbereid en vol vertrouwen aan het onderwijs kon beginnen.

Eline en Colinda, jullie brok aan ervaring en open deuren heb ik altijd immens gewaardeerd. Eline, jij bent altijd bereid om nog een extra taak, een extra meeting of een extra vrijdagmiddagborrel (VrijMiBo) in te plannen. Jij was nooit te flauw om iets voor me uit te zoeken of even met me mee te denken. Ook was je een gewillig slachtoffer voor al mijn VrijMiBo Photoshop kunsten en (onsuccesvolle) airfryer verkooppraatjes, dank! Colinda, ik denk met veel plezier terug aan onze reis naar Boston waar we naast een toffe MPE meeting ook overal ongepland wat later aankwamen omdat we (te) lang hadden gekletst in onze air-BnB. Dit was echter niet verbazingwekkend, want ook op Deb hebben we genoeg lange, maar vooral goede gesprekken gevoerd over allerlei onderwerpen. Bedankt voor al jullie inzichten, ik heb ontzettend veel van jullie geleerd.

(Oud) Epi-aio's, waar moet ik beginnen. Broodjeslunches, Fast-Food Friday, VrijMiBo's, Noro-lunches gevolgd door een heus epidemiologisch onderzoek, journal clubs, rondjes (snel)wandelen, carnaval, congressen en symposia, het is eigenlijk te veel om hier allemaal op te noemen. Maar het belangrijkste bij al deze dingen is dat jullie nooit te flauw waren om de deur wagenwijd open te zetten om te helpen bij vragen, even een praatje te maken, of een keer in die verre uithoek van Deb te kijken of er toch nog iemand zat. Ook al voelt het raar om te vertrekken vanachter een computerscherm, ik waardeer het ontzettend dat we toch nog ZOOM-borrels hebben kunnen organiseren. De laatste borrel in het park maakt het dan nog net iets specialer.

Dan wil ik ook nog even de mensen op 'de serieuze kamer' bedanken voor de geweldige tijd die ik de afgelopen jaren heb gehad. Lloyd en Lisette, wat heb ik het naar mijn zin gehad bij ons op de kamer en ik hoop dat ik jullie de afgelopen jaren niet te veel afgeleid heb. Maar naar mijn idee heb ik daarvoor toch voldoende gecompenseerd door het halen van koffie en thee. Lloyd, als stamhoofd was jij altijd de oude wijsgeer van de kamer. Met jouw methodische en analytische aanpak wist jij altijd snel tot de kern van het probleem te komen, zelfs als dit los stond van wetenschappelijke problemen. Ik denk met veel plezier terug aan alle vrijdagen waar jij bewust bus na bus miste om toch nog even de belangrijke dingen in het leven door te spreken. Mogelijk kan je de banen van de bureaustoelen nog in de vloerbedekking zien staan. Lisette, kamelenkoningin, jij was vaak een bron van inspiratie voor mij als ik jou zag knallen achter die miniscule schermpjes. Als jij je zinnen op iets zet dan gaat het ook gebeuren, maar altijd in een positieve zin. Jij bent denk ik het schoolvoorbeeld van schijn bedriegt, en dan heb ik het niet over imposter-syndroom. Over de jaren bleef je me verbazen met verhalen over whiskeyproeverijen en de meest smerige gangsterrap die ik in eerste instantie nooit bij jou geplaatst zou hebben.

Natuurlijk wil ik ook graag mijn vrienden bedanken voor jullie steun en belangstelling, maar vooral ook voor de (luchtige) ontspanning gedurende de afgelopen vier jaar. Sam en Chris, onze dagen en avondjes gevuld met verhitte discussies over vanalles en nog wat, heerlijke muziek van Vet-Relaxte-Deephouse tot Maasjes-finest en intensieve spellen waren voor mij altijd 'five-out-of-five'! Ook wil ik jullie ontzettend bedanken voor alle momenten waar jullie oprechte interesse toonden in mijn onderzoek met kritische vragen die soms voor mij vanuit

een totaal andere invalshoek kwamen. Hoewel het lijkt dat we in de toekomst een andere aanpak moeten vinden voor onze chillsessies, hoop ik wel dat we deze avondjes gewoon door kunnen zetten. Ik wil daarnaast natuurlijk ook de 'Bison Boys' bedanken. Jullie wisten vaak mijn gedachten te verzetten als het even (te) druk werd of als er belangrijke keuzes gemaakt moesten worden. Hopelijk kunnen we binnenkort weer met zijn allen bij elkaar komen.

Frits, als ware epidemiologiehotline was jij altijd bereid om even te sparren over nieuwe ideeën en toffe analyses, maar ook over dingen zoals de beste wetenschappelijke-SPAM mails van de week of de slechtste epi grappen. Ik vraag me nog geregeld af hoe het bij jou in het Japanse ruimteprogramma verlopen is. Helaas is onze vriendschap nog steeds niet in een co-auteurschap vastgelegd, maar wat niet is, kan nog komen. De toekomst is soms onzekerder dan het breedste betrouwbaarheidsinterval.

Stella en Daan, samen hebben we GW en de HSRM doorlopen. Als ik de goudgele tijden bij M.S.V. Santé vergelijk met de koffiebruine tijden van nu dan kom ik tot de pijnlijke realisatie dat we toch echt wel degelijk geworden zijn. Met alle life-events die er op het moment nog aan zitten te komen zal dat alleen nog maar erger worden. Tsja, het is nog altijd flink wennen om mijn eigen agenda bij te moeten houden, want volgens mij ben ik meestal de bottleneck bij het plannen van onze avondjes! Ik zal alvast wat vrije avondjes inplannen!

(Schoon)familie, bedankt dat jullie altijd vol interesse naar mijn verhalen luisterden. Ook al waren de verhalen en onderwerpen jullie vaak wereldvreemd, toch kon ik rekenen op jullie ongebreidelde aandacht. Als ik er even doorheen zat of als het even tegenzat kon ik ook altijd op jullie geruststelling en steun rekenen. Als ik boos was, waren jullie boos mét mij, als ik blij was, vierden jullie het met me mee, terecht of onterecht, no questions asked. Dank hiervoor, jullie weten maar half wat het voor mij betekent.

Pap en Mam, jullie hebben mij altijd geleerd dat ik mijn dromen na moet jagen, zolang ik maar gelukkig ben. Of dat nu betekent dat ik putjesschepper, minister president of dan toch epidemioloog wil worden, jullie staan onvoorwaardelijk achter mij. Ook zorgen jullie er altijd voor dat ik in het weekend de mogelijkheid heb om mijn hoofd even helemaal leeg te maken in het zand van Budel. Helaas betekende dit vaak dat jullie zelf in de kou stonden te vernikkelen. Sas, Hans en Luka, wat hebben jullie het goed voor elkaar. Hoewel mijn werk vaak maar een vreemde wereld is voor jullie, luisterden jullie altijd naar mijn verhalen als we met zijn allen thuis waren en stelden jullie altijd vragen uit oprechte interesse, bedankt. Natuurlijk hoop ik dat Luka later, mogelijk door een klein beetje stimulans van Ome Jeroen, een succesvol kernfysicus wordt.

Lieve Minke, Baas, Miep, als het in mijn hoofd even te druk wordt dan ben jij er altijd om mij er weer uit te halen. Met jouw zachte karakter, inlevingsgevoel en zorgzaamheid kan ik me dan ook geen betere vriendin wensen. Met plezier denk ik terug aan al onze avondjes rustig op de bank of lachend/schreeuwend (afhankelijk van wie er wint) tijdens al onze bordspelletjes. Samen met jou voelt echt als samen. Ik kan dan ook niet wachten om te zien wat de toekomst ons zal brengen.

Curriculum Vitae

About the author

Jeroen van de Pol was born on October 29th 1990, in Eindhoven, the Netherlands. After graduating from secondary school (Stedelijk College Eindhoven, Eindhoven) in 2010, he studied Health Sciences, with the specialization "Biology & Health" (2011-2014) at Maastricht University. During his bachelor, Jeroen wrote his bachelor's thesis "Effects of heating multiple body parts on thermal comfort and thermal sensation: finding indicators for thermal comfort" in 2014 under supervision from Prof. dr. Wouter D. van Marken Lichtenbelt. Afterwards, he enrolled in the Health Sciences Research Masters' program with the specialization "Clinical Epidemiology" at Maastricht University (2014-2016). He performed his master's internship "What determinants influence microbiota composition and diversity in the gut of children of 6-10 years in the Koala Birth Cohort Study?" under the supervision of Dr. Monique Mommers and Dr. John Penders at the Department of Epidemiology at Maastricht University. After graduating, Jeroen started a PhD project at the Department of Epidemiology at Maastricht University under the supervision of Prof. Dr. Ir. Piet A. van den Brandt, Dr. Leo J. Schouten and Dr. K. Kok (UMCG Groningen). The research presented in this dissertation was conducted within the framework of the Netherlands Cohort Study on diet and cancer. From January 2021, Jeroen wil work as postdoc/project leader at the "Nederlands Hart Netwerk".

List of publications

van de Pol JAA, van Best N, Mbakwa CA, Thijs C, Savelkoul PH, Arts IC, et al. Gut Colonization by Methanogenic Archaea Is Associated with Organic Dairy Consumption in Children. Frontiers in microbiology. 2017;8:355.

van de Pol JAA, van den Brandt PA, Schouten LJ. Kidney stones and the risk of renal cell carcinoma and upper tract urothelial carcinoma: the Netherlands cohort study. British Journal of Cancer. 2019;120:368-374

van de Pol JAA, van den Brandt PA, van Engeland M, Godschalk RWL, Schooten FJ, Hogervorst JGF, Schouten LJ. Gene-environment interactions and germline polymorphisms in the Von Hippel and Hypoxia-inducible factor 1-alphagene and renal cell cancer risk. Scientific Reports. 2020; 10(1):137

van de Pol JAA, George L, van den Brandt PA, Baldewijns MMLL, Schouten LJ. Etiologic heterogeneity of clear-cell and papillary renal cell carcinoma in the Netherlands Cohort Study. International Journal of Cancer. 2020;1-10

Submitted for publication:
van de Pol JAA, van den Brandt PA, Schouten LJ. Type 2 diabetes and its treatment and renal cell cancer risk.

van de Pol JAA, Ferronika P, Westers H, van Engeland M, Terpstra MM, Smits K, de Lange K, van den Brandt PA, Sijmons RH, Schouten LJ, Kok K. Evaluation of a seven-gene mutational profile as a prognostic factor in a population-based study of clear cell renal cell carcinoma.